Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 10 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Commentary

Viewpoints

Corrigenda and Addendas

Research Letters

Original Paper

Mobile Apps for the Management of Gastrointestinal Diseases: Systematic Search and Evaluation Within App Stores

Eva-Maria Messner[1], PhD; Niklas Sturm[2], MD; Yannik Terhorst[1,3], MSc; Lasse B Sander[4], PhD; Dana Schultchen[5], PhD; Alexandra Portenhauser[1], BSc; Simone Schmidbaur[2], MD; Michael Stach[6], MSc; Jochen Klaus[2], MBA, MD, PD; Harald Baumeister[1], PhD, Prof Dr; Benjamin M Walter[2], MD, Prof Dr

[1]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, University of Ulm, Ulm, Germany

[2]Department of Internal Medicine I, University Hospital Ulm, Ulm, Germany

[3]Department of Research Methods, Institute of Psychology and Education, University of Ulm, Ulm, Germany

[4]Department of Rehabilitation Psychology and Psychotherapy, Institute of Psychology, Albert-Ludwigs-University Freiburg, Freiburg at Breisgau, Germany

[5]Department of Clinical and Health Psychology, Institute of Psychology and Education, University of Ulm, Ulm, Germany

[6]Institute of Databases and Information Systems, University of Ulm, Ulm, Germany

Corresponding Author:
Benjamin M Walter, MD, Prof Dr
Department of Internal Medicine I
University Hospital Ulm
Albert-Einstein-Allee 23
Ulm, 89081
Germany
Phone: 49 0731 5000
Email: benjamin.walter@uniklinik-ulm.de

Abstract

Background: Gastrointestinal diseases are associated with substantial cost in health care. In times of the COVID-19 pandemic and further digitalization of gastrointestinal tract health care, mobile health apps could complement routine health care. Many gastrointestinal health care apps are already available in the app stores, but the quality, data protection, and reliability often remain unclear.

Objective: This systematic review aimed to evaluate the quality characteristics as well as the privacy and security measures of mobile health apps for the management of gastrointestinal diseases.

Methods: A web crawler systematically searched for mobile health apps with a focus on gastrointestinal diseases. The identified mobile health apps were evaluated using the Mobile Application Rating Scale (MARS). Furthermore, app characteristics, data protection, and security measures were collected. Classic user star rating was correlated with overall mobile health app quality.

Results: The overall quality of the mobile health apps (N=109) was moderate (mean 2.90, SD 0.52; on a scale ranging from 1 to 5). The quality of the subscales ranged from low (mean 1.89, SD 0.66) to good (mean 4.08, SD 0.57). The security of data transfer was ensured only by 11 (10.1%) mobile health apps. None of the mobile health apps had an evidence base. The user star rating did not correlate with the MARS overall score or with the individual subdimensions of the MARS (all $P>.05$).

Conclusions: Mobile health apps might have a positive impact on diagnosis, therapy, and patient guidance in gastroenterology in the future. We conclude that, to date, data security and proof of efficacy are not yet given in currently available mobile health apps.

(J Med Internet Res 2022;24(10):e37497) doi:10.2196/37497

KEYWORDS

gastrointestinal diseases; mHealth; mobile health; MARS; Mobile Application Rating Scale; systematic review; app quality; gastrointestinal; mobile app; app

Introduction

Gastrointestinal diseases are associated with substantial morbidity and health care costs worldwide [1-5]. For example, in the United States, the annual health care expenditures for gastrointestinal diseases were US $135.9 billion in total, with more than 54.4 million ambulatory visits with a primary diagnosis for gastrointestinal disease and 3.0 million hospital admissions [6]. Additional indirect costs arise due to substantial levels of personal disability, work absenteeism, and loss of productivity [7-12]. Therefore, health care systems are challenged to provide equitable and affordable solutions for patients with digestive diseases [6,13].

In particular, for the successful treatment of chronic gastrointestinal diseases (eg, inflammatory bowel disease [IBD] and irritable bowel syndrome), the patient's adherence and compliance are crucial [14-19]. Treatment recommendations are extensive, consisting of medical and psychological measures [20-24]. Moreover, they include high-demand interventions such as health behavior changes (eg, dietary adjustments or stress management) that cannot be addressed adequately in routine health care [6,24-26]. Additionally, the COVID-19 pandemic with consecutive lockdown forced the health care institutions to uptake contactless approaches [27-32]. Therefore, the implementation of mobile health (mHealth) apps might be a promising approach [33-36].

A recent US study showed that 58.2% of smartphone users had at least 1 mHealth app downloaded on their device [37]. Fitness and nutrition apps were the most commonly downloaded mHealth apps [37]. However, mHealth solutions might also have a potential impact in prevention, diagnostics, and therapy in gastrointestinal disorders [38].

Unfortunately, there is a relevant gap between the high number of available mHealth apps to manage gastrointestinal diseases and the low number of reliable scientific studies in this field [33,36,39]. This gap is concerning as the use of mHealth apps is accompanied with potential risks and side effects such as insufficient data protection and a lack of privacy, as well as treatment without informed consent [40]. Other potential hazards such as misinformation, nonavailability in emergencies, and data misuse have been reported for mHealth apps [40,41].

Due to the rapid development in technology, users and health care providers have difficulties in identifying relevant, high-quality mHealth apps, because they have to rely on the information provided in the stores such as user star ratings and app descriptions [42]. Previous studies have already indicated that user star ratings are potentially misleading because they are influenced by user-friendliness and functionality rather than by content quality [43]. Furthermore, they might be biased due to fake ratings or older versions of the app [42-44]. Therefore, user star ratings might not be a valid orientation aid for selecting a mHealth app, and other strategies to support users and health care providers select an appropriate mHealth app to manage health care issues should be considered.

Additionally, many scientifically tested apps developed by universities and research projects do not enter the app market [45]. In contrast, many available mHealth apps developed by commercial providers have never been tested for their effectiveness and efficacy [45]. Therefore, the quality of publicly available mHealth apps for gastrointestinal diseases is not evident in the literature. Due to increasing public interest in the use of mHealth apps, reliable reviews and analyses are mandatory [46].

Quality-measuring instruments for mHealth apps such as the multidimensional Mobile Application Rating Scale (MARS) are available in several languages, validated, and used worldwide [47-50]. MARS is an expert rating tool that allows researchers to reliably assess and compare mHealth apps regarding user engagement, functionality, aesthetics, and the quality of information [50-52]. Furthermore, it offers a descriptive section in which aims, methods, theoretical background, and cost, etc, can be assessed [48,52]. The MARS was widely used to assess app quality systematically (eg, weight management, rheumatoid arthritis, chronic back pain, mindfulness, heart failure, chronic pain, posttraumatic stress disorder, medication adherence, depression, and smoking cessation, etc) [43,53-59].

The aim of this study was to systematically search for mHealth apps for gastrointestinal diseases in the app stores and evaluate their quality, content, and characteristics using the MARS [48]. Furthermore, mHealth app characteristics such as theoretical background, the content of the apps, affiliation, and price were assessed. Moreover, the accordance with gastroenterological guidelines and evidence base of the included mHealth apps were investigated.

Methods

Study Design

This systematic review was oriented on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [60].

Search Strategy and Procedure

An automatic search engine (Mobile Health App Database [MHAD] web crawler [61]) was used to systematically screen the Google Play and Apple App stores for eligible mHealth apps [62] between October 24, 2020, and June 12, 2021. The applied search terms were defined by conducting focus groups with patients with gastrointestinal disorders and health care providers at the University Hospital Ulm and Freiburg to mimic lay and professional searches. The final search terms included "digestive problems," "stomach pain," "constipation," "CED," "ulcerative colitis," "Crohn's disease," "inflammatory bowel disease," "reflux," "bloating," "diarrhea," "celiac disease," "food intolerances," and "malabsorption." The search terms were entered separately because logical operations and truncation cannot be used in the Google Play and Apple App stores.

All found mHealth apps were registered in a central database, and duplicates were automatically removed. All identified apps were screened regarding whether their title, description, given images, and comments of app users indicated that the app (1) was developed for gastrointestinal health issues, (2) provided

in the German or English language, (3) was downloadable in the official Google Play or Apple App store, (4) was functional to enable an assessment (no device problems), and (5) met no other exclusion criteria (app bundles, only usable with another device such as a smartwatch, or not active for download). In a second step, the apps were downloaded and checked regarding the aforementioned criteria.

Data Extraction, Evaluation Criteria, and Instruments

The included apps were evaluated by raters using the German version of the multidimensional MARS (MARS-G) [48]. Before starting with the evaluation process, the raters received standardized web-based training, which is publicly accessible and free of charge [63]. For quality assurance, interrater reliability (IRR) between the 2 raters was calculated. Rater agreement was examined by intraclass correlation (ICC) based on a 2-way mixed-effect model. A minimum ICC of .75 was predefined as sufficient ICC [64]. An additional reviewer was consulted when the IRR was below a value of .75 [48,64].

Evaluation Tool MARS-G

The evaluation tool MARS-G is a reliable and valid procedure for the quality assessment of mHealth apps [48,52]. The MARS-G has a very good internal consistency for overall score (ω=.82, 95% CI 0.76-0.86) and high levels of IRR (2-way mixed ICC=.84, 95% CI 0.82-0.85) [48].

General Characteristics

For examining app characteristics, the classification page of the MARS-G was used. It contains (1) the app name; (2) app version; (3) platform; (4) content-related subcategory; (5) store link; (6) price; (7) user star rating; (8) the number of user star ratings; (9) theoretical background (eg, type of therapy); (10) aims; (11) methods (eg, information/education, monitoring and tracking, gamification, and reminder); (12) technical aspects (eg, allows sharing); (13) data protection and safety (eg, password protection); (14) field of application; and (15) certification [48,50]. The classification site of MARS-G was used to assess the content and functions of the included mHealth apps [50,59]. With the MARS-G, a descriptive assessment of privacy and security features is possible. All features were assessed based on the information included in the mHealth apps or app stores. External information was not evaluated.

Quality Assessment

The multidimensional quality rating of the MARS-G consists of 6 different subdimensions with 19 items, which can be evaluated on a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent): (1) engagement (entertainment, interest, individual adaptability, interactivity, and target group); (2) functionality (performance, usability, navigation, motor, and gestural design); (3) aesthetics (layout, graphics, and visual appeal); and (4) information (accuracy of app description, goals, quality and quantity of information, quality of visual information, credibility, and evidence base); (5) subjective quality (recommendable, probability of using the app in the next 12 months, payment, and star rating); and (6) perceived impact (increased awareness, increased knowledge, attitudes, fosters intention to change, empowers help-seeking behavior, and fosters behavior change) [48,50]. For the assessment of the overall quality, the total score was calculated from the 4 main subdimensions (engagement, functionality, aesthetics, and information) [50]. The ratings of the reviewers were averaged for all calculations. Mean scores and SDs were calculated for the MARS overall score and subdimensions.

Quality Rating on Evidence

To verify whether empirical studies were available for the mHealth apps, item 19 on the information subscale of the MARS was used. This item was examined by searching the mHealth apps' name in Google, Google Scholar, PubMed, and the developers or providers' website for existing efficacy and effectiveness studies [48].

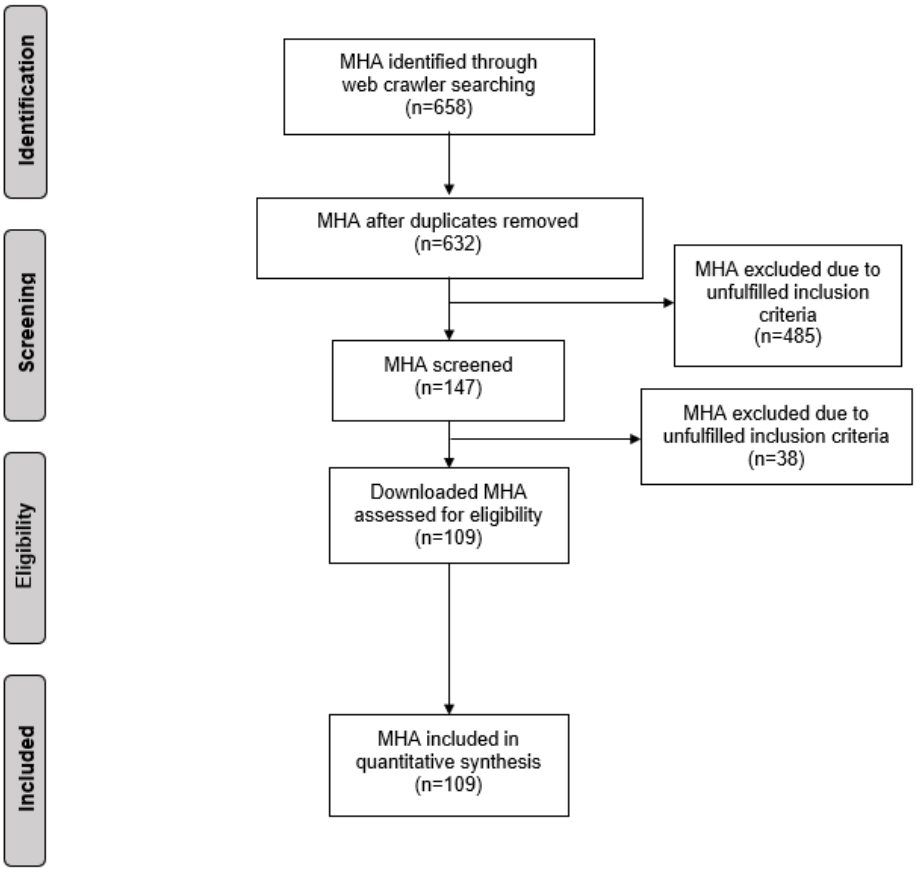
User Star Rating

The user star ratings were extracted from the app stores. The user star rating from Google Play and Apple App stores is rated on a scale of 1 to 5 stars. It is presented as a cumulative average of individual ratings in the app stores [65]. Pearson correlation coefficient between user star ratings and MARS-G ratings were calculated. For all analysis, an α level of 5% was defined [66].

Results

The web crawler identified 658 mHealth apps, of which 109 were eligible for inclusion after screening and eligibility check (Figure 1).

Figure 1. Flowchart of the inclusion process of mobile health apps (MHA).



General Characteristics

Of the 109 mHealth apps, 79 (72.5%) were from the Google Play store, and 30 (27.5%) were from the Apple App store; 53 (48.6%) had a user star rating, whereas 56 (51.4%) were not rated by store users. The mean user star rating was 3.96 (SD 0.80), ranging from 2.00 to 5.00.

Most apps (n=93, 85.3%) were free of charge, and the prices of fee-based mHealth apps ranged from €0.69 to €8.99 (mean €4.0, SD €2.25; from US $0.84 to US $10.91; mean US $4.86, SD US $2.73). The 109 mHealth apps for gastrointestinal disorders were identified in the following Google Play or Apple App store categories (multiple categories can be assigned to 1 mHealth app): “health and fitness” (n=76, 69.7%); “medical” (n=33, 30.3%); “food and drinks” (n=11, 10.1%); “lifestyle” (n=3, 2.8%); “books and references” (n=2, 1.8%); “education” (n=2, 1.8%); “entertainment” (n=3, 2.8%); and “parenting” (n=1, (0.9%; Table 1).

The included mHealth apps targeted the following aims (multiple aims may be selected for 1 mHealth app): “improvement of general well-being” (n=92, 84.4%); “promotion of physical health” (n=86, 78.9%); “entertainment” (n=3, 2.8%); “support for behavioral changes” (n=33, 30.3%); “support in achieving individual goals” (n=27, 24.8%), “reduction of stress” (n=7, 6.4%); “reduction of fear” (n=4, 3.7%), “improvement of social behavior” (n=2, 1.8%); and “other aims” (n=16, 14.7%)—for example, “information” (n=3, 2.8%) or “education” (n=2, 1.8%; Table 2).

Table 1. Frequency of the app store categories of the mobile health apps for gastrointestinal disorders (multiple selection possible).

App store category	App (N=109), n (%)
Parenting	1 (0.9)
Medical	33 (30.3)
Lifestyle	3 (2.8)
Health and fitness	76 (69.7)
Food and drinks	11 (10.1)
Entertainment	3 (2.8)
Education	2 (1.8)
Books and references	2 (1.8)

Table 2. Frequency of the aims of the mobile health apps for gastrointestinal disorders (multiple selection possible).

Aim	App (N=109), n (%)
Improvement of general well-being	92 (84.4)
Promotion of physical health	86 (78.9)
Entertainment	3 (2.8)
Support for behavioral changes	33 (30.3)
Support in achieving individual goals	27 (24.8)
Reduction of stress	7 (6.4)
Reduction of fear	4 (3.7)
Improvement of social behavior	2 (1.8)
Other aims	16 (14.7)

Content and Functions

Of the 109 mHealth apps, almost all (n=91, 83.5%) focused on educational information about gastrointestinal diseases; over half (n=71, 65.1%) offered specific "tips and advice"; and the following methods were also frequent: "monitoring and tracking" (n=22, 20.2%), "alternative medical intervention elements" (n=18, 16.5%), "data collection and measurement" (n=13, 11.9%), feedback (n=13, 11.9%), and "memory, reminder, and amplifier" (n=7, 6.4%). The frequency of the methods used is summarized in Table 3.

Almost all mHealth apps (n=101, 92.7%) had "treatment" as their field of application. Other frequent fields were "prevention of disease" (n=73, 67%), "rehabilitation" (n=51, 46.8%), and "aftercare" (n=45, 41.3%).

Table 3. Frequency of methods in the included mobile health apps for gastrointestinal disorders (multiple selection possible).

Method	App (N=109), n (%)
Information and education	91 (83.5)
Tips and advice	71 (65.1)
Monitoring and tracking	22 (20.2)
Alternative intervention elements	18 (16.5)
Data collection and measurement	13 (11.9)
Feedback	13 (11.9)
Memory, reminder, and amplifier	7 (6.4)
Pursuing own goals	5 (4.6)
Traditional medicine	4 (3.7)
Strategies, skills, and training	2 (1.8)
Relaxing exercises	2 (1.8)
Gamification	2 (1.8)
Tailored interventions and real-time feedback	2 (1.8)
Other	1 (0.9)
Physical exercises	1 (0.9)
Mindfulness and gratefulness	1 (0.9)
Acceptance	1 (0.9)

Privacy and Security Features

Of the 109 mHealth apps, 9 (8.2%) had no privacy and security features; 69 (63.3%) had an imprint, and 54 (49.5%) had a visible privacy policy; 16 (14.7%) required consent to data collection in an active form, and 54 (49.5%) in a passive form; and 11 (10.1%) ensured the security of data transfer, 11 (10.1%) required a log-in, 13 (11.9%) offered a password protection system, 7 (6.4%) informed about the conflicts of interests or financial background, and 1 (0.9%) had an emergency function.

Quality Rating

The overall quality of mobile health apps was average (mean 2.90, SD 0.52; ranging from 1.84 to 4.47). The top 10 ranked mHealth apps with the highest overall quality are listed in Tables 4 and 5. Concordance between raters was good to excellent (ICC from 0.76, 95%CI 0.70-0.81 to 0.93, 95% CI 0.92-0.94).

The average quality ratings of all included mHealth apps of the MARS subscales were the following: engagement, 2.47 (SD 0.74; range 1.10-5.00); functionality, 4.08 (SD 0.57; range 2.25-5.00); aesthetics, 3.19 (SD 0.76; range 1.17-4.83); and information quality, 1.89 (SD 0.66; range 0.57-3.79). The subjective quality was 2.16 (SD 0.79; range 1.00-4.50) and the perceived impact was 2.33 (SD 0.63; range 1.15-4.08; Table 6).

Table 4. Top 10 ranked mobile health apps according to Mobile Application Rating Scale overall quality, target, developer, and category.

App	Rating, mean	Target[a]	Developer	Category[b]
vyoapp - Die CED-App	4.47	Digestive problems	Takeda Pharma Vertriebs GMbH & Co. KG	Medical
My IBD Manager from AGA	4.18	Ulcerative colitis	@Point of care	Health and fitness
MyColitis	4.05	Ulcerative colitis	myColitis	Health and fitness
My IBD Care	3.86	Inflammatory bowel disease	Ampersand health limited	Medicine and health and fitness
Cliexa-IBD	3.85	Inflammatory bowel disease	CN4CE, Inc	Medical
Poop Tracker – Toilet Login	3.82	Digestive problems	Appstronaut Studios	Health and fitness
Doc4Me – IBD Doctor Search	3.71	Inflammatory bowel disease	The North American Society for Pediatric Gastroenterology, Hepatology and Nutrition and Gotomo GmbH	Medical and health and fitness
Food Navi – Coeliac	3.71	Celiac disease	Goe GmbH	Health and fitness and food and drink
Histamin, Fructose & Co.	3.69	Food intolerance	Baliza GmbH	Health and fitness and food and drink
Reflux Tracker	3.68	Digestive problems	Gotomo GmbH	Health and fitness

[a]Target disease or search term.

[b]Category in the Apple App or Google Play store.

Table 5. Privacy policy, informed consent, certification, and price of the top 10 ranked mobile health apps.

App	Privacy policy[a]	Informed consent[b]	Certification[c]	Price, € (US $)
vyoapp - Die CED-App	Yes	No	No	0 (0)
My IBD Manager from AGA	Yes	Yes	American Gastroenterological Association	0 (0)
MyColitis	Yes	No	No	0 (0)
My IBD Care	Yes	Yes	No	0 (0)
Cliexa-IBD	Yes	Yes	No	0 (0)
Poop Tracker – Toilet Login	Yes	No	No	0 (0)
Doc4Me – IBD Doctor Search	Yes	Yes	No	0 (0)
Food Navi – Coeliac	No	No	No	3.49 (4.24)
Histamin, Fructose & Co.	Yes	No	No	5.99 (7.27)
Reflux Tracker	No	No	No	0 (0)

[a]Mobile health app had a privacy policy that could be accessed.

[b]Informed consent was actively obtained.

[c]Mobile health app was certified or developed under professional surveillance.

Table 6. Subjective quality rating and the rating of perceived impact on user according to the Mobile Application Rating Scale.

Variable	Rating, mean (SD)
Subjective quality rating	2.35 (0.84)
Recommendable	2.17 (0.94)
Probability of using the app in the next 12 months	2.53 (1.06)
Payment	1.31 (0.58)
Star rating	2.63 (0.89)
Perceived impact	2.31 (0.64)
Increased awareness	2.46 (0.90)
Increased knowledge	2.60 (1.00)
Attitudes	2.14 (0.65)
Fosters intention to change	2.10 (0.83)
Empowers help-seeking behavior	2.22 (1.17)
Fosters behavior change	2.49 (0.83)

Quality Rating on Evidence

Only 2 (1.8%) of the 109 mHealth apps were certified and developed in concordance with guidelines published by the American Gastroenterological Association. None of the mHealth apps had an evidence base.

Correlation Patterns

The user star rating did not correlate with the MARS overall score or the individual subdimensions (overall: r=–0.03; P=.86; engagement: r=–0.11; P=.46; functionality: r=–0.17; P=.23; aesthetics: r=0.15; P=.28; information: r=0.02; P=.87; subjective quality: r=0.07; P=.61; perceived impact: r=–0.12; P=.39).

Discussion

Principal Findings

This study is the first that comprehensively and systematically reviewed mHealth apps for different gastrointestinal disorders available in the Google Play and Apple App stores [39]. The quality of the mHealth apps was investigated by standardized expert ratings using the MARS-G [48]. In total, 109 mHealth apps with a focus on gastrointestinal disorders were included. Therefore, this analysis offers the first comprehensive systematic expert review of mHealth apps in the field of gastroenterology.

The majority of the mHealth apps were found in the categories “health and fitness” and “medical.” The average quality of the included apps was moderate, according to the applied quality criteria. Only 2 mHealth apps were certified and developed in concordance with approved guidelines such as those from the American Gastroenterological Association. This fact is alarming because the concordance of a mHealth app with approved guidelines is crucial to prevent mistreatment and misinformation. A similar lack of adherence to well-established medical guidelines was found in mHealth app quality reviews for depression and posttraumatic stress disorder [57,62]. Moreover, our data show that user star ratings did not correlate with the experts’ MARS ratings. However, this finding is in accordance with a previous study on mHealth apps for posttraumatic stress disorder and in contradiction to a systematic review of mHealth apps for mindfulness [59,67]. These findings underline the need for systematic reviews to empower patients and health care providers in informed health care decisions. Freely available platforms, which display expert quality ratings of mHealth apps such as the MHAD [61], Psyberguide [68], or KVAppradar [69], have been installed as a possible solution to empower patients and health care providers. In addition to these platforms that offer an evaluation of available mHealth apps based on the general criteria of scientific evidence, professional gastroenterological societies should participate in the development and assessment of mHealth apps in consideration of established guidelines. Regarding the rapid progress in the methods of disease monitoring and therapy of gastrointestinal disorders, suitable apps should be constantly updated for adequate support. In particular, for long-term gastrointestinal disorders, such as IBD, which are characterized by an unstable disease course with recurrent remission and exacerbation, mHealth apps could be a promising approach for symptom monitoring with an early detection of disease relapse. As previous studies have shown that self-reporting symptom diaries correlate with disease activity index for Crohn disease [70,71], validated symptom assessment questionnaires could be implemented in future mHealth apps.

From the patients and health care providers’ perspectives, mHealth interventions could demonstrate a great potential to facilitate the monitoring of symptoms, improve self-management–related physical or psychosocial consequences, and maintain compliance [72-77]. Rapid advancement in mobile technology may enable real-time data capture and exchange between patient self-monitoring devices and a remote monitoring system, which creates promising opportunities to provide prompt feedback to patient-generated alerts and specific needs [38].

Besides the lack of mHealth apps for adequate symptom monitoring, our results showed that none of the evaluated apps were designed to evaluate adverse drug reactions that occur during disease therapy. Giraud et al [78] have demonstrated that

40.9% (N=1179) of patients with IBD that participated in the IBDREAM registry had at least 1 adverse drug reaction, and 24 new adverse drug reactions were found based on their analysis. These findings suggest that the evaluation of adverse events during maintenance therapy in IBD and possibly other gastrointestinal diseases should be monitored closely to timely change or adapt drug dose or substance choice for individual-tailored therapy. The use of mHealth apps for the monitoring of adverse drug reactions, especially during the start of a new therapeutical agent, could be a new field for the implementation of mHealth apps in clinical practice. The clinical monitoring of disease activity and drug compatibility could be further enhanced by wearable devices that track physical parameters and by noninvasive biomarker monitoring (eg, c-reactive protein or interleukin-1 for IBD from sweat [79]). In their comprehensive review, Chong and Woo [80] have demonstrated that approaches for the implementation of wearable sensor systems for gastrointestinal disease already exist and could change clinical practice in the near future [80].

Furthermore, the results highlight the need for a comprehensive evaluation of clinical effectiveness and economic effects. In particular, the long-term effects and cost-effectiveness of mHealth apps to manage gastrointestinal diseases should be elaborated in future studies [38]. Currently, studies that have evaluated the cost-effectiveness of telemedicine-directed treatment and monitoring of IBD show a reduction of hospitalization and therapy costs [81] but remain controversial regarding the total cost-effectiveness of telemedical interventions [82]. Since the use of biologicals has been identified as the major cost driver for IBD [83], mHealth apps could help to early de-escalate and optimize biological treatment after constant disease remission and enhance conventional therapy admission to prevent unnecessary therapy escalation to expensive biologicals. To date (2022), in Germany, a central register for Conformité Européenne–certified eHealth apps with scientifically proven benefit for patients has been established. Apps that are listed in the national digitale Gesundheitsanwendungen (digital health care app in German) register are prescriptible by health care professionals. The costs could be reimbursed by the patient's health insurance companies, which might be a step toward the implementation of trustworthy, certified apps into daily health care.

Additionally, the review revealed that data security is not always guaranteed when using mHealth apps. As health data are highly sensitive, this lack of guarantee is one reason why the use of mHealth apps in the management of gastrointestinal diseases cannot be clearly recommended currently. We found that the security of data transfer was only ensured in 14% of the mHealth apps. As patient safety is paramount, data security is a keystone for adopting mobile technologies into health care. In this field, respect for privacy, security, the disclosure of data sharing, traceability, and the guarantee of transparency are essential factors. These factors are in line with other reviews of the data security and privacy of mHealth apps for smoking cessation, depression, and older adults [40,41,46].

When using and implementing mHealth technologies into health care systems, it will be important to know how these technologies will fit within the existing organizational framework, which may involve changes in business structure and culture, workflow, and staff. In this context, the primarily legal aspects of mHealth app use play a substantial role. National regulations for mHealth approaches such as the act on medical devices—the Medical Devices Directive—for the European Union or the Food and Drug Administration regulation body for the United States exist. The harmonization of the regulation instruments is crucial for the sufficient uptake of mHealth solutions worldwide. Such worldwide standards for the safe use of mHealth apps in gastroenterology should include (1) being based on current standards and medical guidelines, (2) randomized controlled trial testing for effectiveness, (3) high standards for data security, and (4) minimal and economic data recording.

We acknowledge several limitations regarding this review. First, due to the rapid growth and dynamic changes in mHealth apps available on the global market, this study can only represent a snapshot view of the available mHealth apps as of July 2021 for the management of gastrointestinal disorders. The continuous monitoring of the market is mandatory to reliably inform users and health care providers. Second, the main focus was on English- and German-language medical mHealth apps, which might have impaired the generalizability of the results, as the quality of mHealth apps may vary between countries and continents. Third, the review included all types of gastrointestinal disorders with a focus on inflammatory and nutritive bowel diseases. An even more precise analysis of mHealth apps addressing the multiple subspecialties of gastrointestinal disorders could be promising. Furthermore, the analysis of mHealth apps for hepatobiliary disease and gastrointestinal cancer (eg, mHealth apps for the patient-related surveillance of adverse events due to chemotherapy) should be evaluated specifically in further studies. Fourth, the user star ratings in the app stores may refer to various versions of an mHealth app and are aggregated across the different versions. Therefore, the MARS rating and the user star rating could refer to different versions.

Conclusion

This systematic review of mHealth apps that manage gastrointestinal diseases found a moderate overall quality of mHealth apps available in app stores. The quality of user engagement and information quality was rated as poor, thus limiting the possible positive effects of mHealth app use to manage gastrointestinal diseases. Furthermore, data safety and privacy were mostly not given. Moreover, there were no efficacy studies on the included mHealth apps, and only 2 mHealth apps were following well-established guidelines for the treatment of gastrointestinal diseases. Taken together, these findings implicate a red flag of the use of currently available mHealth apps for the management of gastrointestinal diseases. Nevertheless, given the possible positive impact of mHealth apps in the routine care of individuals with gastrointestinal diseases, an improvement in the quality of medical content for mHealth apps and data safety is mandatory.

Acknowledgments

The authors would like to the thank Jiaxi Lin, Rüdiger Pryss, Robin Kraft, Pascal Damasch, and Philipp Dörzenbach for their support in the development of the search engine and their support in the Mobile Health App Database [MHAD] project. We would also like to thank Linda Armbruster for her assistance in the rating of the mobile health apps.

Data Availability

The primary data of the systematic review can be provided by the corresponding author on a reasonable request. Data will only be shared for scientific purposes. Data sharing agreements may have to be signed depending on the request.

Authors' Contributions

EMM, YT, LBS, BMW, and HB developed the study design. DS, LBS, AP, JK, and MS collected the data. AP and SS conducted the statistical evaluations. EMM, AP, and BMW wrote the first draft of the paper. All authors contributed to the current version of the paper and approved the final paper. EMM is the guarantor of the paper.

Conflicts of Interest

HB codeveloped and run the German Mobile Health App Database (MHAD) project. The MHAD is a self-funded project at Ulm University with no commercial interests. HB, LBS, and EMM received payments for talks and workshops as well as license fees in the context of e-mental health without a specific link to the mobile health apps rated in this paper. All other authors declare no other conflicts of interest.

References

1. Beard JA, Franco DL, Click BH. The burden of cost in inflammatory bowel disease: a medical economic perspective and the future of value-based care. Curr Gastroenterol Rep 2020 Jan 30;22(2):6. [doi: 10.1007/s11894-020-0744-z] [Medline: 32002671]
2. Silverstein MD, Loftus EV, Sandborn WJ, Tremaine WJ, Feagan BG, Nietert PJ, et al. Clinical course and costs of care for Crohn's disease: Markov model analysis of a population-based cohort. Gastroenterology 1999 Jul;117(1):49-57. [doi: 10.1016/s0016-5085(99)70549-4] [Medline: 10381909]
3. Cohen RD, Yu AP, Wu EQ, Xie J, Mulani PM, Chao J. Systematic review: the costs of ulcerative colitis in western countries. Aliment Pharmacol Ther 2010 Apr;31(7):693-707 [FREE Full text] [doi: 10.1111/j.1365-2036.2010.04234.x] [Medline: 20064142]
4. Sperber AD, Bangdiwala SI, Drossman DA, Ghoshal UC, Simren M, Tack J, et al. Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome Foundation Global Study. Gastroenterology 2021 Jan;160(1):99-114.e3 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.014] [Medline: 32294476]
5. Oka P, Parr H, Barberio B, Black CJ, Savarino EV, Ford AC. Global prevalence of irritable bowel syndrome according to Rome III or IV criteria: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2020 Oct;5(10):908-917. [doi: 10.1016/S2468-1253(20)30217-X] [Medline: 32702295]
6. Peery AF, Crockett SD, Murphy CC, Lund JL, Dellon ES, Williams JL, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. Gastroenterology 2019 Jan;156(1):254-272.e11 [FREE Full text] [doi: 10.1053/j.gastro.2018.08.063] [Medline: 30315778]
7. Seyedian SS, Nokhostin F, Malamir MD. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. J Med Life 2019;12(2):113-122 [FREE Full text] [doi: 10.25122/jml-2018-0075] [Medline: 31406511]
8. Kawalec P. Indirect costs of inflammatory bowel diseases: Crohn's disease and ulcerative colitis. a systematic review. Arch Med Sci 2016 Apr 01;12(2):295-302 [FREE Full text] [doi: 10.5114/aoms.2016.59254] [Medline: 27186172]
9. Constantin J, Atanasov P, Wirth D, Borsi A. Indirect costs associated with ulcerative colitis: a systematic literature review of real-world data. BMC Gastroenterol 2019 Nov 09;19(1):179 [FREE Full text] [doi: 10.1186/s12876-019-1095-9] [Medline: 31706270]
10. Kuenzig ME, Lee L, El-Matary W, Weizman AV, Benchimol EI, Kaplan GG, et al. The impact of inflammatory bowel disease in Canada 2018: indirect costs of IBD care. J Can Assoc Gastroenterol 2019 Feb;2(Suppl 1):S34-S41 [FREE Full text] [doi: 10.1093/jcag/gwy050] [Medline: 31294383]
11. Buono JL, Carson RT, Flores NM. Health-related quality of life, work productivity, and indirect costs among patients with irritable bowel syndrome with diarrhea. Health Qual Life Outcomes 2017 Feb 14;15(1):35 [FREE Full text] [doi: 10.1186/s12955-017-0611-2] [Medline: 28196491]
12. Frändemark Å, Törnblom H, Jakobsson S, Simrén M. Work productivity and activity impairment in irritable bowel syndrome (IBS): a multifaceted problem. Am J Gastroenterol 2018 Oct;113(10):1540-1549. [doi: 10.1038/s41395-018-0262-x] [Medline: 30254230]

13. Park KT, Colletti RB, Rubin DT, Sharma BK, Thompson A, Krueger A. Health insurance paid costs and drivers of costs for patients with Crohn's disease in the United States. Am J Gastroenterol 2016 Jan;111(1):15-23. [doi: 10.1038/ajg.2015.207] [Medline: 26195179]
14. Herman ML, Kane SV. Treatment nonadherence in inflammatory bowel disease: identification, scope, and management strategies. Inflamm Bowel Dis 2015 Dec;21(12):2979-2984. [doi: 10.1097/MIB.0000000000000581] [Medline: 26355470]
15. Tabibian A, Tabibian JH, Beckman LJ, Raffals LL, Papadakis KA, Kane SV. Predictors of health-related quality of life and adherence in Crohn's disease and ulcerative colitis: implications for clinical management. Dig Dis Sci 2015 May 6;60(5):1366-1374. [doi: 10.1007/s10620-014-3471-1] [Medline: 25559754]
16. Roditi D, Robinson ME. The role of psychological interventions in the management of patients with chronic pain. Psychol Res Behav Manag 2011;4:41-49 [FREE Full text] [doi: 10.2147/PRBM.S15375] [Medline: 22114534]
17. Chan W, Chen A, Tiao D, Selinger C, Leong R. Medication adherence in inflammatory bowel disease. Intest Res 2017 Oct;15(4):434-445 [FREE Full text] [doi: 10.5217/ir.2017.15.4.434] [Medline: 29142511]
18. Lenti MV, Selinger CP. Medication non-adherence in adult patients affected by inflammatory bowel disease: a critical review and update of the determining factors, consequences and possible interventions. Expert Rev Gastroenterol Hepatol 2017 Mar;11(3):215-226. [doi: 10.1080/17474124.2017.1284587] [Medline: 28099821]
19. Bager P, Julsgaard M, Vestergaard T, Christensen LA, Dahlerup JF. Adherence and quality of care in IBD. Scand J Gastroenterol 2016 Nov;51(11):1326-1331. [doi: 10.1080/00365521.2016.1195870] [Medline: 27311071]
20. Feingold J, Murray HB, Keefer L. Recent advances in cognitive behavioral therapy For digestive disorders and the role of applied positive psychology across the spectrum of GI care. J Clin Gastroenterol 2019 Aug;53(7):477-485. [doi: 10.1097/MCG.0000000000001234] [Medline: 31169757]
21. Montero AM, Jones S. Roles and impact of psychologists in interdisciplinary gastroenterology care. Clin Gastroenterol Hepatol 2020 Feb;18(2):290-293. [doi: 10.1016/j.cgh.2019.07.067] [Medline: 31401360]
22. Basnayake C, Kamm MA, Stanley A, Wilson-O'Brien A, Burrell K, Lees-Trinca I, et al. Standard gastroenterologist versus multidisciplinary treatment for functional gastrointestinal disorders (MANTRA): an open-label, single-centre, randomised controlled trial. Lancet Gastroenterol Hepatol 2020 Oct;5(10):890-899. [doi: 10.1016/S2468-1253(20)30215-6] [Medline: 32679040]
23. Black CJ, Thakur ER, Houghton LA, Quigley EMM, Moayyedi P, Ford AC. Efficacy of psychological therapies for irritable bowel syndrome: systematic review and network meta-analysis. Gut 2020 Aug;69(8):1441-1451. [doi: 10.1136/gutjnl-2020-321191] [Medline: 32276950]
24. Black CJ, Drossman DA, Talley NJ, Ruddy J, Ford AC. Functional gastrointestinal disorders: advances in understanding and management. Lancet 2020 Nov 21;396(10263):1664-1674. [doi: 10.1016/S0140-6736(20)32115-2] [Medline: 33049221]
25. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan 17;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
26. Palsson OS, Ballou S. Hypnosis and cognitive behavioral therapies for the management of gastrointestinal disorders. Curr Gastroenterol Rep 2020 Jun 03;22(7):31. [doi: 10.1007/s11894-020-00769-z] [Medline: 32495233]
27. Spinelli A, Pellino G. COVID-19 pandemic: perspectives on an unfolding crisis. Br J Surg 2020 Jun;107(7):785-787 [FREE Full text] [doi: 10.1002/bjs.11627] [Medline: 32191340]
28. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
29. Bokolo AJ. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
30. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. Aten Primaria 2020;52(6):418-422 [FREE Full text] [doi: 10.1016/j.aprim.2020.04.003] [Medline: 32402477]
31. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the future. Telemed J E Health 2020 May;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
32. Ananthakrishnan AN, Singh S. The doctor will call you now! telemedicine in the midst of a pandemic. Clin Gastroenterol Hepatol 2020 Jul;18(8):1688-1690 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.031] [Medline: 32315781]
33. Kelso M, Feagins L. Can smartphones help deliver smarter care for patients with inflammatory bowel disease? Inflamm Bowel Dis 2018 Jun 08;24(7):1453-1459. [doi: 10.1093/ibd/izy162] [Medline: 29868764]
34. Helsel BC, Williams JE, Lawson K, Liang J, Markowitz J. Telemedicine and mobile health technology are effective in the management of digestive diseases: a systematic review. Dig Dis Sci 2018 Jun 16;63(6):1392-1408. [doi: 10.1007/s10620-018-5054-z] [Medline: 29663265]
35. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: scoping review. J Med Internet Res 2019 Aug 19;21(8):e14630 [FREE Full text] [doi: 10.2196/14630] [Medline: 31429410]

36. Kernebeck S, Busse TS, Böttcher MD, Weitz J, Ehlers J, Bork U. Impact of mobile health and medical applications on clinical practice in gastroenterology. World J Gastroenterol 2020 Aug 07;26(29):4182-4197 [FREE Full text] [doi: 10.3748/wjg.v26.i29.4182] [Medline: 32848328]
37. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
38. Riaz MS, Atreja A. Personalized technologies in chronic gastrointestinal disorders: self-monitoring and remote sensor technologies. Clin Gastroenterol Hepatol 2016 Dec;14(12):1697-1705 [FREE Full text] [doi: 10.1016/j.cgh.2016.05.009] [Medline: 27189911]
39. Con D, de Cruz P. Mobile phone apps for inflammatory bowel disease self-management: a systematic assessment of content and tools. JMIR mHealth uHealth 2016 Feb 01;4(1):e13 [FREE Full text] [doi: 10.2196/mhealth.4874] [Medline: 26831935]
40. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
41. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. BMJ 2019 Mar 20;364:l920 [FREE Full text] [doi: 10.1136/bmj.l920] [Medline: 30894349]
42. Armstrong S. Which app should I use? BMJ 2015 Sep 09;351:h4597. [doi: 10.1136/bmj.h4597] [Medline: 26353800]
43. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
44. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
45. Rathner EM, Probst T. Mobile Applikationen in der psychotherapeutischen Praxis: Chancen und Grenzen. Mobile applications in psychotherapeutic practice: opportunities and risks. Psychotherapie im Dialog 2018 Nov 28;19(04):51-55. [doi: 10.1055/a-0592-0496]
46. Portenhauser AA, Terhorst Y, Schultchen D, Sander LB, Denkinger MD, Stach M, et al. Mobile apps for older adults: systematic search and evaluation within online stores. JMIR Aging 2021 Feb 19;4(1):e23313 [FREE Full text] [doi: 10.2196/23313] [Medline: 33605884]
47. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016 Jul 07;16(1):83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
48. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR mHealth uHealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
49. Payo MR, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov S, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
50. 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]
51. 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]
52. 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 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
53. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for people with rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. JMIR mHealth uHealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
54. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
55. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR mHealth uHealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
56. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the Mobile App Rating Scale. JMIR mHealth uHealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]

57. Terhorst Y, Rathner E, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. 'Help from the app store?': a systematic review of depression apps in German app stores. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
58. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
59. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' a systematic review and standardised rating of apps for post-traumatic stress disorder (PTSD). Eur J Psychotraumatol 2020 Jan 09;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
60. 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):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
61. Mobile Health App Database. MHAD. URL: http://mhad.science/ [accessed 2022-08-25]
62. Stach M, Kraft R, Probst T, Messner E, Terhorst Y, Baumeister H, et al. Mobile health app database - a repository for quality ratings of mHealth apps. 2020 Sep 01 Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN p. 427-432. [doi: 10.1109/CBMS49503.2020.00087]
63. MARS - Mobile Anwendungen Rating Skala. MARS-Rating Universität Ulm YouTube page. 2019 Jul 31. URL: https://www.youtube.com/watch?v=5vwMiCWC0Sc [accessed 2022-08-25]
64. Fleiss JL. Design and Analysis of Clinical Experiments. Hoboken, NJ: John Wiley & Sons; 2011.
65. Mojica Ruiz IJ, Nagappan M, Adams B, Berger T, Dienst S, Hassan AE. Examining the rating system used in mobile-app stores. IEEE Softw 2016 Nov;33(6):86-92. [doi: 10.1109/ms.2015.56]
66. Zar JH. Significance testing of the Spearman rank correlation coefficient. J Am Stat Assoc 1972 Sep;67(339):578-580. [doi: 10.1080/01621459.1972.10481251]
67. Schultchen D, Terhorst Y, Holderied T, Stach M, Messner E, Baumeister H, et al. Stay present with your phone: a systematic review and standardized rating of mindfulness apps in European app stores. Int J Behav Med 2021 Oct 20;28(5):552-560 [FREE Full text] [doi: 10.1007/s12529-020-09944-y] [Medline: 33215348]
68. Apps and digital health resources reviewed by experts. One Mind PsyberGuide. URL: https://onemindpsyberguide.org [accessed 2022-08-25]
69. Informationen zu Gesundheits-Apps. KVAppradar. URL: https://www.kvappradar.de [accessed 2022-08-25]
70. Kim ES, Park KS, Cho KB, Kim KO, Jang BI, Kim EY, Daegu–Gyeongbuk Gastrointestinal Study Group (DGSG). Development of a web-based, self-reporting symptom diary for Crohn's Disease, and its correlation with the Crohn's Disease Activity Index. J Crohns Colitis 2017 Dec 04;11(12):1449-1455. [doi: 10.1016/j.crohns.2014.09.003] [Medline: 25246007]
71. Echarri A, Vera I, Ollero V, Arajol C, Riestra S, Robledo P, et al. The Harvey-Bradshaw index adapted to a mobile application compared with in-clinic assessment: the MediCrohn study. Telemed J E Health 2020 Jan;26(1):80-88 [FREE Full text] [doi: 10.1089/tmj.2018.0264] [Medline: 30848700]
72. Aitken M, Lyle M. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2015.
73. Endl R, Jäschke T, Thiel C, Wickinghoff DV. mHealth im Kontext des elektronischen Patientendossiersine Studie im Auftrag von eHealth Suisse. mHealth in the context of the electronic patient record: a study commissioned by eHealth Suisse. Fachhochschule St Gall. 2015 Mar 19. URL: https://e-health-com.de/fileadmin/user_upload/dateien/Downloads/Studie_mHealth_Maerz_2015.pdf [accessed 2022-08-25]
74. Kuhn B, Amelung V. Kapitel 4. Gesundheits-Apps und besondere Herausforderungen. Chapter 4. health apps and special challenges. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Hannover, Germany: Medizinische Hochschule Hannover; 2016:100-114.
75. Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res 2015 Feb 12;17(2):e26 [FREE Full text] [doi: 10.2196/jmir.3757] [Medline: 25679446]
76. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
77. Dicianno BE, Parmanto B, Fairman AD, Crytzer TM, Yu DX, Pramana G, et al. Perspectives on the evolution of mobile (mHealth) technologies and application to rehabilitation. Phys Ther 2015 Mar;95(3):397-405 [FREE Full text] [doi: 10.2522/ptj.20130534] [Medline: 24925075]
78. Giraud EL, Thomas PWA, van Lint JA, van Puijenbroek EP, Römkens TEH, West RL, IBDREAM registry. Adverse drug reactions from real-world data in inflammatory bowel disease patients in the IBDREAM Registry. Drug Saf 2021 May;44(5):581-588 [FREE Full text] [doi: 10.1007/s40264-021-01045-3] [Medline: 33538994]
79. Jagannath B, Lin K, Pali M, Sankhala D, Muthukumar S, Prasad S. A sweat-based wearable enabling technology for real-time monitoring of IL-1β and CRP as potential markers for inflammatory bowel disease. Inflamm Bowel Dis 2020 Sep 18;26(10):1533-1542. [doi: 10.1093/ibd/izaa191] [Medline: 32720974]

80. Chong KPL, Woo BKP. Emerging wearable technology applications in gastroenterology: a review of the literature. World J Gastroenterol 2021 Mar 28;27(12):1149-1160 [FREE Full text] [doi: 10.3748/wjg.v27.i12.1149] [Medline: 33828391]
81. de Jong MJ, Boonen A, van der Meulen-de Jong AE, Romberg-Camps MJ, van Bodegraven AA, Mahmmod N, et al. Cost-effectiveness of telemedicine-directed specialized vs standard care for patients with inflammatory bowel diseases in a randomized trial. Clin Gastroenterol Hepatol 2020 Jul;18(8):1744-1752. [doi: 10.1016/j.cgh.2020.04.038] [Medline: 32335133]
82. Ankersen DV, Weimers P, Marker D, Teglgaard Peters-Lehm C, Bennedsen M, Rosager Hansen M, et al. Costs of electronic health vs. standard care management of inflammatory bowel disease across three years of follow-up-a Danish register-based study. Scand J Gastroenterol 2021 May 28;56(5):520-529. [doi: 10.1080/00365521.2021.1892176] [Medline: 33645378]
83. Park KT, Ehrlich OG, Allen JI, Meadows P, Szigethy EM, Henrichsen K, et al. Inflamm Bowel Dis 2020 Jan 01;26(1):1-10 [FREE Full text] [doi: 10.1093/ibd/izz104] [Medline: 31112238]

Abbreviations

IBD: inflammatory bowel disease
ICC: intraclass correlation
IRR: interrater reliability
MARS: Mobile Application Rating Scale
MARS-G: German version of the Mobile Application Rating Scale
MHAD: Mobile Health App Database
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 23.02.22; peer-reviewed by U Bork, L Birrell; comments to author 17.03.22; revised version received 30.05.22; accepted 29.07.22; published 05.10.22.

Please cite as:
Messner EM, Sturm N, Terhorst Y, Sander LB, Schultchen D, Portenhauser A, Schmidbaur S, Stach M, Klaus J, Baumeister H, Walter BM
Mobile Apps for the Management of Gastrointestinal Diseases: Systematic Search and Evaluation Within App Stores
J Med Internet Res 2022;24(10):e37497
URL: https://www.jmir.org/2022/10/e37497
doi:10.2196/37497
PMID:36197717

Review

Artificial Intelligence Applications in Health Care Practice: Scoping Review

Malvika Sharma[1], MSc, MD; Carl Savage[1,2], PhD; Monika Nair[2], PhD; Ingrid Larsson[2], PhD; Petra Svedberg[2], PhD; Jens M Nygren[2], PhD

[1]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Medical Management Centre, Stockholm, Sweden
[2]School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Jens M Nygren, PhD
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 30118
Sweden
Phone: 46 035167100
Email: jens.nygren@hh.se

Abstract

Background: Artificial intelligence (AI) is often heralded as a potential disruptor that will transform the practice of medicine. The amount of data collected and available in health care, coupled with advances in computational power, has contributed to advances in AI and an exponential growth of publications. However, the development of AI applications does not guarantee their adoption into routine practice. There is a risk that despite the resources invested, benefits for patients, staff, and society will not be realized if AI implementation is not better understood.

Objective: The aim of this study was to explore how the implementation of AI in health care practice has been described and researched in the literature by answering 3 questions: What are the characteristics of research on implementation of AI in practice? What types and applications of AI systems are described? What characteristics of the implementation process for AI systems are discernible?

Methods: A scoping review was conducted of MEDLINE (PubMed), Scopus, Web of Science, CINAHL, and PsycINFO databases to identify empirical studies of AI implementation in health care since 2011, in addition to snowball sampling of selected reference lists. Using Rayyan software, we screened titles and abstracts and selected full-text articles. Data from the included articles were charted and summarized.

Results: Of the 9218 records retrieved, 45 (0.49%) articles were included. The articles cover diverse clinical settings and disciplines; most (32/45, 71%) were published recently, were from high-income countries (33/45, 73%), and were intended for care providers (25/45, 56%). AI systems are predominantly intended for clinical care, particularly clinical care pertaining to patient-provider encounters. More than half (24/45, 53%) possess no action autonomy but rather support human decision-making. The focus of most research was on establishing the effectiveness of interventions (16/45, 35%) or related to technical and computational aspects of AI systems (11/45, 24%). Focus on the specifics of implementation processes does not yet seem to be a priority in research, and the use of frameworks to guide implementation is rare.

Conclusions: Our current empirical knowledge derives from implementations of AI systems with low action autonomy and approaches common to implementations of other types of information systems. To develop a specific and empirically based implementation framework, further research is needed on the more disruptive types of AI systems being implemented in routine care and on aspects unique to AI implementation in health care, such as building trust, addressing transparency issues, developing explainable and interpretable solutions, and addressing ethical concerns around privacy and data protection.

(J Med Internet Res 2022;24(10):e40238) doi:10.2196/40238

KEYWORDS

artificial intelligence; health care; implementation; scoping review; technology adoption

Introduction

Artificial intelligence (AI) is often heralded as a potential disruptor that will transform the practice of medicine [1,2]. The promise of AI lies in its ability to process and learn from large volumes of data and capture patterns otherwise difficult for humans to identify. This ability has raised questions and worries about liability and risks, in particular related to the level of autonomy granted to AI applications [3]. Others see a role complementary to humans; for example, decision support or decision augmentation where humans (in the roles of clinicians or programmers) provide oversight and collaborate [4-7]. The latter approach has been demonstrated to yield superior performance compared with experts alone [8]. Other benefits include improved patient outcomes, error reduction, health system optimization, cost reductions, and increased value [6].

The amount of data collected and available in health care, coupled with advances in computational power, has contributed to advances in AI applications [9] and an exponential growth of publications on AI in health care, with >10,000 records on PubMed in 2021 alone. Included in this are multiple reviews across medical specialties that explore the potential roles of AI to augment health care delivery [10-14]. These include diagnostic (eg, early cancer diagnosis, diabetes retinopathy screening, or COVID-19 diagnosis based on computed tomography images), therapeutic (eg, precision medicine in chemotherapy and for combination drug therapy), and regulatory or administrative applications (eg, coding of records or economic evaluations), as well as for population health management (eg, public health surveillance or predictive epidemiological modeling) [15-21].

However, the development of AI applications does not guarantee their adoption into routine health care practice. Research has identified a number of factors influencing adoption of innovations. These include context (eg, economic and political context, laws and regulations, and sociocultural factors), organization (eg, organizational structure, resources, and processes), group (eg, professional values and cultures), individual (eg, attitudes, motivation, user satisfaction, and trust), and technology (eg, usability, design, accuracy, and explainability) [22,23]. This suggests a need to know more about how AI can be implemented in health care, not only as an innovation but also with respect to its unique potential and associated concerns.

Previous reviews have tended to focus only on some aspects of the process of implementation of AI in health care; for example, regulation and legal issues [24,25], trust and ethics [24-29], clinical and patient outcomes [30-32], and economic impact [33]. Others have focused their studies on specific AI applications for health care, such as predictive medicine, diagnostics, and clinical decision-making [9,30,34,35]. A few reviews have been more overarching, focusing on coproduction processes [36], implementation frameworks [37], and critical implementation barriers or success factors [38] that could inform the development of relevant implementation strategies of AI technology. Generally, it is argued that the implementation of AI in health care could significantly improve patient and health care outcomes, but none of these reviews have actually explored the knowledge base of real-world implementation in everyday clinical practice.

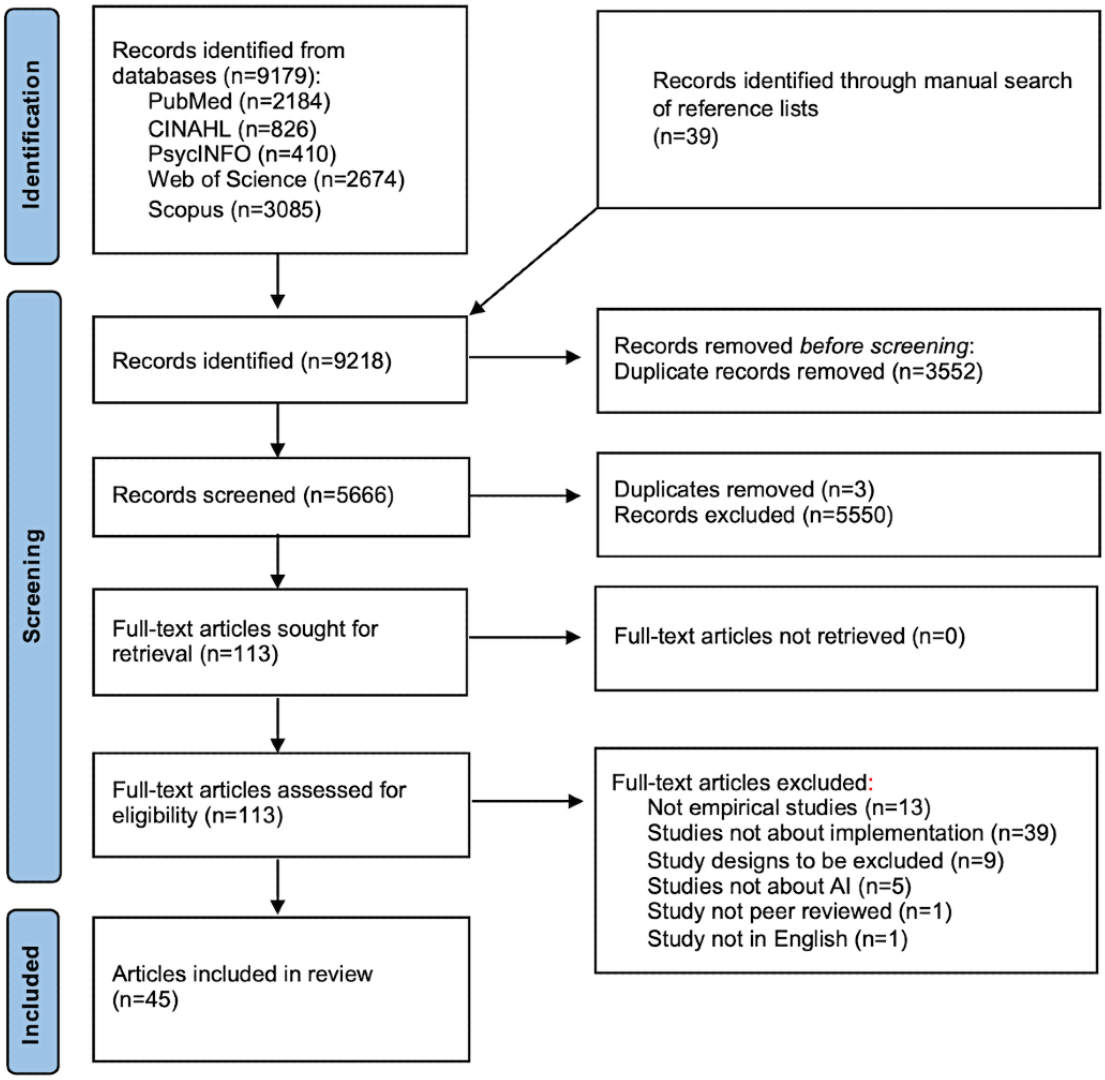
Given the resources invested in developing AI applications and the risk of reproducing already investigated aspects of effective AI applications to support, augment, and perhaps even transform health care for patients, staff, and society, we sought to explore how the implementation of AI in health care practice has been empirically investigated in the research literature.

Methods

Study Design

We chose a scoping review methodology in line with the Arksey and O'Malley framework [39] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Figure 1) [40]. A previous review suggested that implementation of AI in health care was not well studied [37]. A scoping review would thus enable a mapping of the "extent, range and nature of research activity" in this emerging area of research [39].

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



Identifying the Research Question

To address our aim, we formulated three research questions:

1. What are the characteristics of research on implementation of AI in practice?
2. What types and applications of AI systems are described?
3. What characteristics of the implementation process for AI systems are discernible?

Identifying Relevant Studies

We focused our search, with support from a university librarian, by iteratively testing synonyms for 3 concepts: *artificial intelligence*, *health care*, and *implementation* (Textbox 1). For the purposes of clarity, we differentiated between AI algorithms and models (the actual code), AI applications (the innovation *package*), and AI systems (the application in its context) and used standardized Medical Subject Headings terms and subject headings describing AI and its subcategories provided by the databases used for our searches [41]. Implementation was defined as "An intentional effort designed to change or adapt or uptake interventions into routines," based on a review of frameworks for the translation of AI into health care practice [37]. Synonyms were joined by the Boolean operator *OR*; next, we combined the search strings for each concept with the Boolean operator *AND* (Multimedia Appendix 1).

To cover content in both general and health- and health care–specific sources, 5 electronic databases were searched: MEDLINE (PubMed), Scopus, Web of Science, CINAHL, and PsycINFO. In addition, we used snowball sampling by manually reviewing reference lists of the review articles we had identified during the screening that might contain relevant references given the topic of the review.

Textbox 1. Concept areas and synonyms used to develop the search strategy.

Search concepts, combined using "AND"

- Artificial Intelligence
- Healthcare
- Implementation

Search terms, combined using "OR"

- Artificial intelligence, Neural networks, Deep learning, Machine learning
- Delivery of healthcare, Health care, Healthcare
- Implementation, Improvement, Innovation, Intervention

Eligibility Criteria

We included peer-reviewed empirical studies published in English between December 2011 and February 2022 because preliminary searches suggested that AI applications in health care are a more recent phenomenon (Table 1).

Table 1. Eligibility criteria and their rationale.

Eligibility criteria and variable	Rationale
Inclusion criteria	
Peer reviewed	Greater credibility because the papers have been reviewed by peer experts in the field
Empirical study design	Empirical studies improve the ability to answer the research questions compared with conceptual commentaries or viewpoints
Published between December 2011 and February 2022	Given the rapid pace of development of technology and changing data sets, solutions developed before the last decade are likely to be obsolete
English language	Practical consideration, given the investigators' language proficiency
Exclusion criteria	
Nonempirical designs, including editorials, commentaries, opinion articles, and reports	Empirical studies improve the ability to answer the research questions compared with conceptual commentaries or viewpoints
Proof-of-concept, feasibility, or validation studies not related to implementation of artificial intelligence technologies	As the aim was to explore implementation in practice, studies that stop short of that, for example, proof-of-concept, validity, or feasibility studies, should be excluded

Study Selection

All identified records were imported into the open-access software Rayyan. Duplicates were removed, and the titles and abstracts of the remaining records were screened for eligibility by at least one of the authors. Any uncertainty or conflict was discussed at regular check-ins until consensus was reached among all authors. These discussions were informed by the multidisciplinary backgrounds of the authors. We also continually reviewed our interpretations of the screening criteria, and when questions were raised, we backtracked to ensure that the criteria had been applied correctly and in a universal fashion, independent of who had screened the records. We used the AI screening and highlighting function of Rayyan, but we still screened each record. We also erred on the side of inclusion. Full-text articles were then screened independently by at least two researchers. Conflicts and uncertainty were again resolved through discussion until consensus was reached among all researchers. As we followed the original framework, a quality appraisal of the included studies was not conducted.

Charting the Data

We developed a data extraction template to chart data for each of the research questions. To define these conceptual areas, we adopted the World Health Organization's guidance on ethics and governance of AI for health definition of AI (based on a recommendation of the Council on Artificial Intelligence of the Organisation for Economic Co-operation and Development states) [42,43]: "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy" [42].

The following data were extracted:

1. General information: authors, publication year, country, clinical setting, study aim, and study design
2. Types and applications of AI: AI technology used, type of AI model, type of task performed by AI, level of action autonomy, intended use of AI, and intended user of AI

3. Implementation process: research focus, motives for implementation, elements in the implementation process, and frameworks used

Collating, Summarizing, and Reporting the Results

The extracted data relating to research questions 1 and 2 were mapped and summarized. A qualitative thematic analysis [44] was used to analyze data associated with research question 3 to summarize the motives for implementation and elements in the implementation process. Articles were read and reread, with initial ideas sorted into either the domain *Motives* behind the implementation or *Elements* in the implementation process. Next, initial codes were identified in each article. The codes were compared based on similarities and differences and collated into potential themes, which were then compared to generate a thematic map that was used to generate clear definitions and names for each theme in the respective domains. Coding and data analysis were performed in pairs, and any uncertainties were discussed among all authors until consensus was achieved.

Results

Search Results

We identified 9218 records, of which 9179 (99.58%) were identified through database searches and 39 (0.04%) through a snowball search of reference lists in the review articles (n=36). Of the 9218 records, after removal of duplicates, 5666 (61.47%) records remained, and we screened titles and abstracts. In this screening, 98% (5553/5666) of the records were excluded, and the remaining 2% (113/5666) were assessed for eligibility through full-text review. Of these 113 articles, 68 (60.2%) were excluded for reasons highlighted in Figure 1, and 45 (39.8%) were included in the scoping review.

Research Question 1: Study Characteristics

The reviewed body of literature was fairly recent, with the majority of the studies (32/45, 71%) having been published between 2020 and 2022 [45-76]. Most (33/45, 73%) of the articles were from North America and Europe [46,47,49-55,57,58,61-63,67-70,73-87], of which most (18/33, 55%) were from the United States [46,47,49-52,54,68,73-77,79-81,84,87]. The greatest number of AI systems were implemented either in hospital-wide settings (6/45, 13%) [50,55,56,65,74,80] or in radiology (6/45, 13%) [53,56,66,68,73,76]. Most (27/45, 60%) of the studies were authored by a multidisciplinary team [46,47,50-55,58,59,61,62,64,67,69,70,72,74,75,78-80,82,86-89], with clinical and IT or informatics backgrounds being the most common combination (9/27, 33%) [47,50,55,61,70,74,79,87,89]. Among studies with authors from only 1 domain, the most common background was clinical (8/45, 18%) [63,65,66,68,71,73,76,84]. There was a wide range of study designs. Most (24/45, 53%) used a case-study design, including both single-case [46,49,50,52,53,55-57,59,60,66-68,70,74,75,78-83,85,86] or multiple comparative case designs [53,56,78] (Table 2 and Multimedia Appendix 2).

Table 2. Overview of articles included in the scoping review (N=45).

Author, year, country; clinical setting	Study aim	Study design
Anand et al [79], 2018, United States; pediatrics	Describe Child Health Improvement through Computer Automation system and methods to represent pediatric guidelines using Arden syntax	Case study
Baxter et al [50], 2020, United States; hospital-wide implementation	Conduct a detailed analysis of barriers to use of machine learning model in health care	Case study
Bennet [77], 2011, United States; mental health	Evaluate the effects of a data-driven clinical productivity system that leverages electronic health record data to provide productivity decision support functionality in a real-world clinical setting	Pre-post study
Champion et al [87], 2011, United States; intensive care	Illuminate barriers and facilitators to use of intensive insulin therapy CDSS[a]	Qualitative study
Chonde et al [68], 2021, United States; radiology	Evaluate the implementation of an AI[b]-powered translation system in radiology	Case study
Chong et al [65], 2021, Australia; hospital-wide implementation	Determine if a VTE[c] stewardship program can increase risk-appropriate VTE prophylaxis and VTE risk assessment using CDSS	Interrupted time series
Cruz et al [85], 2019, Spain; primary care	Describe a real-time CDSS and its effect on adherence to clinical pathways	Case study
Damoah et al [60], 2021, Ghana; management	Explore how an AI-enhanced medical drone application in Ghana's health care supply chain improves the health care supply chain system	Case study
Davis et al [73], 2020, United States; radiology	Determine the impact of a machine learning algorithm, meant to mark CT[d] head examinations pending interpretation as higher probability for intracranial hemorrhage	Case study
Dios et al [83], 2015, Spain; surgery	Present a decision support system for operating room scheduling at a university hospital in Seville, Spain	Case study
García Bermúdez et al [69], 2021, Spain; internal medicine service	Assess the user satisfaction of a virtual caregiver designed to monitor the health of patients admitted to hospital for COVID-19 infection for a period of 30 days after discharge	Quantitative study
Goncalves et al [59], 2020, Brazil; nursing	Present the nurses' experience with technological tools to support the early identification of sepsis	Case study
Herman et al [64], 2021, Indonesia; public health	Assess the impact of an AI-based application on rifampicin-resistant tuberculosis screening	Qualitative study with key informant interviews
Kalil et al [88], 2018, Brazil; surgery	Describe the impact of a new risk-management cognitive robot related to the processes of identification and care for patients at sepsis risk in a clinical-surgical unit	Retrospective observational study
Kashyap et al [47], 2021, United States; not specified	Identify the different computational and organizational setups that early-adopter health systems have used to integrate an AI-based CDSS into clinical workflows	Qualitative study with key informant interviews
Lacey et al [61], 2020, United Kingdom; surgery	Assess the impact of using automatic video auditing in the quality and quantity of hand-wash events	Interrupted time series
Lai et al [52], 2020, United States; public health	Describe the implementation of a digitally automated prehospital triage solution to direct patients to appropriate care	Case study
Litvin et al [84], 2012, United States; primary care	Describe use of a CDSS on antibiotic prescribing for acute respiratory infections in primary care, as well as facilitators and barriers to adoption	Mixed methods
McKillop et al [48], 2021, multiple regions; public health	Characterize the diverse use cases of COVID-19–related conversational agents built using the IBM Watson Assistant platform	Cross-sectional study
Mohamed et al [71], 2021, United Arab Emirates; dentistry	Validate and implement the AI system and quantify referral patterns to the orthodontist specialist before and after implementation of the system	Quantitative survey
Moorman [49], 2021, United States; inpatient care	Describe the experiences and lessons learned during implementation of AI system	Case study
Morales et al [72], 2021, Brazil; emergency care	Describe early implementation of a digital triage and monitoring service that included the use of a chatbot using algorithmic decision-making	Observational study
Ng et al [45], 2021, Singapore; general care	Develop a predictive model for risk stratification for enrollment into a nationwide transitional care program	Analysis of existing data set

Author, year, country; clinical setting	Study aim	Study design
O'Neil et al [76], 2021, United States; radiology	Assess (1) whether the introduction of an algorithm for the detection of intracerebral hemorrhage at noncontrast CT affects turnaround times and (2) whether the impact on turnaround time was dependent on the manner in which information was presented in the radiologist workflow	Quasi-experimental study
Petitgand et al [67], 2020, Canada; emergency department	Analyze the implementation of an AI-based decision support system in an emergency department focusing on actors' representations of the system	Case study
Rais et al [82], 2018, Portugal; management	Discuss optimization approaches for logistics services in hospitals	Case study
Rath et al [81], 2017, United States; surgery	Describe the development, implementation, and evaluation of a model-based decision support system to determine daily scheduling of anesthesiologists and rooms for elective surgeries	Case study
Reis et al [55], 2020, Germany; hospital-wide implementation	Describe a failed AI project at a large hospital and identify the root causes that led to failure	Case study
Romero-Brufau et al [51], 2020, United States; primary care	To explore attitudes about AI among staff who used AI-based CDSS	Pre-post study
Romero-Brufau et al [54], 2020, United States; general care units	Reduce unplanned hospital readmissions using AI-based CDSS	Controlled study
Saverino et al [62], 2021, Italy; rehabilitation	Describe the role of a digital AI platform in facilitating the implementation of changes in rehabilitation service during the COVID-19 pandemic	Retrospective observational study
Schlicher et al [75], 2021, United States; management	Discuss the implementation of data analytics in AI-enabled mission control at one of the largest health care service providers in Washington state	Case study
Schuh et al [78], 2018, Austria; intensive care, oncology, and nephrology	Outline the technical and clinical aspects of 3 CDSSs integrated into practice at Vienna General Hospital	Case study describing 3 projects
Semenov et al [86], 2016, Russia; laboratory	Present research and development of a decision support system for the patients of a laboratory service	Case study
Sendak et al [46], 2020, United States; emergency department	Describe the steps taken to integrate Sepsis Watch, a sepsis detection and management platform, into routine care delivery at Duke University Hospital in Durham, North Carolina	Case study
Snowdon et al [74], 2020, United States; interdisciplinary	Describe the system implemented, workflow changes, and impact on vulnerable citizens	Case study
Strohm et al [53], 2020, The Netherlands; radiology	Identify barriers and facilitators to the implementation of AI applications in clinical radiology	Case study (multiple)
Sukums et al [89], 2015, Ghana and Tanzania; primary care	Describe health workers' acceptance and use of the CDSS for maternal care at rural facilities in Ghana and Tanzania and identify factors affecting successful adoption	Mixed methods
Sun [56], 2021, China; hospital-wide implementation	Study how social power among various stakeholders affects IT adoption in health care	Mixed methods
Tamposis et al [70], 2022, Greece; urology	Present design and implementation of a software platform for supporting detection as well as using and processing clinical, bio-chemical, imaging, and histopathologic findings from fusion biopsy	Case study
Tan et al [66], 2021, Singapore; radiology	Describe the use of AI for automatic detection and flagging of CT findings not reported by radiologists to improve patient safety	Case study
Thurso et al [58], 2021, Slovakia; dentistry	Evaluate the clinical impact of an AI upgrade of an existing orthodontic mobile coaching app	Pre-post study
Wen et al [80], 2019, United States; hospital-wide implementation	Present recommendations for developing natural language processing tool sets based on the experience of developing clinical natural language processing at the Mayo Clinic in Rochester, Minnesota	Case study
Wijnhoven [57], 2021, The Netherlands; neonatal care	Theory formalization of grounded insights from a CDSS development case, and by doing this create an organizational learning theoretical foundation for AI development in organizations	Case study

Author, year, country; clinical setting	Study aim	Study design
Wong et al [63], 2021, Canada; oncology	Characterize the impact of deep learning–based auto-segmented contour models in the clinical workflow at 2 cancer centers	User feedback survey

[a]CDSS: clinical decision support system.

[b]AI: artificial intelligence.

[c]VTE: venous thromboembolism.

[d]CT: computed tomography.

Research Question 2: Types and Applications of AI Technology

The most common type of AI application implemented was automation or optimization technology, reported in 71% (32/45) of the implemented systems [45,46,49-51,53-59,62,64,65,70,71,73,75,77-79,81-84,86-89]. Other technologies implemented included human language technologies, computer vision, and robotics technology (Table 2 and Multimedia Appendices 2 and 3). The most common AI model was a symbolic or knowledge-based model, reported in nearly half (22/45, 49%) of the reviewed studies [48,52-54,57,59,68-74,77-80,84,85,88], followed by statistical models (9/45, 20%) [45,49-51,58,81,82]. The most commonly performed task was recognition (16/45, 36%) [52,56,61,63-66,72,73,76,78-80,84,85], followed by forecasting (9/45, 20%) [45,46,49-51,53,54,57,71]. Other tasks performed were event detection, goal-driven optimization, interaction support, and personalization (Table 2, Multimedia Appendices 2 and 3). Although more than half (24/45, 53%) of the AI applications had no action autonomy [46,48-51,53,54,57,63,66,67,70,73-75,79,81-85,87-89], a few reported applications had low (2/21, 10%) [55,72], medium (4/21, 19%) [58,69,71,86], or high (6/21, 29%) [52,55,60,61,68,76] action autonomy (Table 2, Multimedia Appendices 2 and 3). Nearly three-quarters of all AI systems were intended for clinical care (33/45, 73%) [46,49,51,53-59,61,63-73,78-80,84-89], and the majority (18/33, 55%) of these concerned providing support to inform the patient-provider encounter [46,49,51,55,56,61,63,65,67,68,74,78,79,84,85,87,89], followed by diagnosis and prediction-based diagnosis (13/33, 39%) [53,55,57,59,64,66,70,71,73,78,80,86,88]. The remaining AI systems (12/45, 27%) were intended for health systems management and planning [45,50,52,60,62,74-77,81-83]. Health care providers were the most common target users; most often physicians (19/45, 42%) [46,49,51,53-55,57-64,66-68,70,71,73,74,76,79,80,84,85,88,89], followed by nurses (6/45, 13%) [46,49,51,59,87,88]. Other intended users included health workers, technicians, managers, patients or caregivers, and the general public (Table 2 and Multimedia Appendix 2).

Research Question 3: Implementation Process Characteristics

The research focus in approximately a third of the studies was to present the effectiveness of the implemented intervention (16/45, 36%) [54,58,60-62,65,66,71,73-75,77,81,82,85,88]. Other research foci included user experiences [51,59,63,64,69,86], AI use metrics [48,52,80,84,89], and identification of barriers or facilitators [50,53,55,57,67,87] (Table 2, Multimedia Appendices 2 and 3). Most (32/45, 71%) of the studies described the implementation process as successful, and only a few (4/45, 9%) described it as unsuccessful (in the rest of the studies, the success of the implementation was either not mentioned, or the outcome was inconclusive).

In a little more than half (23/45, 51%) of the reviewed studies, the motives behind the implementation were not described. For those studies that did (22/45, 49%), we identified 6 types of motives, with *Improve health care quality* and *Achieve better patient outcomes* being the 2 most common. Studies in the former theme described AI systems used to improve quality of services [46,71,75,87,88], reduce diagnostic errors [66], reduce hospital length of stay [73], or reduce unplanned readmissions [50,54], whereas studies in the latter theme described AI systems used to achieve better patient survival [59,70]. Another theme, *Improve efficiency*, focused on health care–cost reduction, increased service production, and optimization of public services [45,72,74,76,77]. *Respond to the COVID-19 pandemic* was stated as a motive necessitated by the need for access to the most up-to-date information [48], the sudden surge in demand for health care services [52], prioritization of limited resources [72], and reorganization of service delivery in response to local guidelines for prevention of infection transmission [62]. *Improve provider satisfaction* focused on workload reduction for health care professionals [55,69]. *Empower patients* by using AI to support interpretations of laboratory investigations, rather than just the test results, was another motive for implementing AI [86].

Of the 45 included studies, 3 (7%) had an explicit focus on implementation processes [46,49,68]. In the other studies, characteristics common to implementation processes were identified: cocreation, contextualization, nondisruptive workflow design, communication, learning focus, training, incentives, and organizational strategies. Both barriers and facilitators were described.

Several (8/45, 18%) implementation efforts involved *cocreation* with multidisciplinary stakeholders, starting from an ideation phase that included problem identification, requirement collection, and design or redesign of clinical workflows to facilitate AI-system integration [45,46,49,52,55,59,68,78]. Cocreation also involved end users in the design of user interfaces [46,68]. *Contextualization* of AI systems relating to the local context and target population was highlighted as important in development and implementation [52,54]. *Nondisruptive workflow design* was emphasized, where efforts were made to design AI systems around existing roles and functions of the intended user to avoid radical modification of current practice to fit the AI system [46,49,51]. *Communication* efforts were seen as central to building trust and promoting use

by sharing evidence of AI effectiveness with clinicians and describing overall benefits of the technology [46,49,59], appointing champions to promote AI among peers [46,53,75], and encouraging informal communication between clinicians and IT developers to cultivate relationships and build trust in the AI [56]. However, the study by Sendak et al [46] encouraged the separation of developers and clinicians and made conscious efforts to shift focus away from the technical aspects of AI. A *learning focus* could begin in the ideation phase to understand and assess the problem to be addressed by AI before coding, through development and implementation, by iteratively testing and adjusting workflows [46]. After implementation, learning continued through the continuous capture of user feedback to enable improvement [68]. *Training* involved both informal and formal sessions to enable AI use [56,89]. After implementation, training could continue in formal peer-group meetings to share best practices and individual training and support for more reluctant users [84]. *Incentives* were used to promote or enforce AI use. More controlling approaches included periodic monitoring and audits [56,84] or removing alternative ways of performing the task altogether to necessitate AI use [84]. Gamification was used to promote a feeling of reward and competition [61,65]. *Organizational efforts* involved including the hospital's top leadership as essential members of the project team and the design and implementation of the AI system to promote uptake [49,55]. One organization formed a special governance committee as a formal mechanism to monitor AI use among health care providers [46]. Another organization's innovation strategy included innovation managers as part of the organizational structure to promote AI [53].

In 7% (3/45) of the studies [50,57,68], the use of the following implementation frameworks was mentioned: the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework [90]; the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework [91]; and the Socialization, Externalization, Combination, and Internalization model of knowledge dimensions [92]. Of the 45 included studies, 4 (9%) proposed new frameworks, principles, or recommendations based on their presented findings and implementation experiences [49,55,56,80]. Moorman [49] proposed 6 principles for implementation of AI: elements of trust and transparency, minimal impact on workflows, stakeholder buy-in, relevant education, actionability of AI outputs, and sustainability through follow-up interactions. Reis et al [55] proposed a framework for overcoming cognitive and affective resistance to AI implementation centered around concerns of users (physicians), such as transparency and understandability of the AI system, involvement of users in the AI training, and trust in the AI system. Sun [56] proposed a power strategy matrix for AI adoption, suggesting that a "boss strategy" or "expert strategy" can influence adoption. Wen et al [80] presented 3 desiderata for developing an AI-based platform, where the second one focused on improving adoption.

Discussion

Principal Findings

Our aim with this study was to explore how the implementation of AI in health care practice has been empirically investigated in the research literature. We found that research on implementation of AI systems is mostly published in high-income countries, covers many different clinical settings and disciplines, and predominantly focuses on care providers as users. The AI models are primarily symbolic or knowledge based, use automation or optimization technologies, and are mainly used to perform tasks related to recognition. AI systems are predominantly intended for clinical care, particularly clinical care pertaining to patient-provider encounters. Most possess no action autonomy but rather support human decision-making. The focus of most research is on establishing the effectiveness of interventions or related to technical and computational aspects of AI systems. Focus on the specifics of implementation processes does not yet seem to be a priority in research, and the use of frameworks to guide implementation is rare.

Study Characteristics

Most of the studies were published very recently (2020-2022), which is unsurprising given the temporal distribution of AI health care studies. Research on AI implementation in health care is predominantly conceptual in nature, dominated by commentaries, perspectives, opinion articles, and conceptual frameworks that raise important questions and issues but without much-needed empirical evidence [93-96]. As the empirical evidence base for the implementation of AI solutions in routine health care is still narrow and premature, it limits possibilities for generalization both for practice and for the advancement of methodological approaches. Most of the articles were published in high-income countries, particularly the United States. This finding is consistent with the more developed digital health infrastructure, routine use of electronic health records, and big data initiatives in North American and European countries and aligns with other reviews of AI applications in various fields of health care [32,97,98]. The many different clinical settings and disciplines could corroborate the data-driven nature of health care; the fact that AI is highly applicable; or that because of its nascent state, AI is still being tried in many different contexts. Given the focus on clinical care, it is not surprising that the intended users were mostly health care providers, particularly physicians. A recent scoping review on the use of AI in primary care found a similar predominance of physicians as target end users [99]. This suggests a view of AI systems as tools to support decision-making by physicians rather than other health professionals. It was surprising to find a scarcity of implementations of AI applications to handle infectious diseases (except for the study by McKillop et al [48]), given the overwhelming attention given to, and funding provided for, the management of the COVID-19 pandemic in 2020-2022. Another underrepresented area where AI holds a strong promise is mental health (except for the studies by Bennett [77] and Rahman et al [100]).

Types and Applications of AI Technology

Nearly half of the AI models were symbolic or knowledge based. They used human-generated logical representations, rules, and ontologies to infer conclusions and have greater explainability than models that are based on pure data-driven or statistical approaches. However, they might not live up to the full potential of AI because they are "hard-coded, expert cookbooks" that are limited by the knowledge that is encoded into them [101]. Data-driven, statistical approaches such as machine learning learn predictive functions based on the inputted data. However, these methods are opaque and have implications for health care in relation to patient or provider trust, accountability and quality assurance, and patient safety [3,102]. The World Health Organization's guidance on ethics and governance of AI for health recognizes the potential trade-off between transparency and accuracy but encourages AI explainability and transparency over black-box approaches [43]. The predominance of knowledge-based or symbolic models, whose greater transparency and longer existence may ease acceptance among care providers, is in line with previous reviews [103]. However, the majority of recently published AI models use data-driven or hybrid technologies, and knowledge-based models comprised only a minority of the applications [104]. Our study found that automation or optimization technologies were by far the most common, followed by human language technologies. More than half of the AI systems implemented had no action autonomy. Instead, they were human decision support systems where the AI system cannot act on its recommendation or output but depends on the human operating the system to use or disregard the recommendation made by it. This finding indicates that decision support systems are the types of AI systems that have achieved adoption the earliest, likely because they enhance human actions and cause minimal disruption to clinical workflows [105].

Implementation Process

This study found that the way the implementation process of AI systems in health care is researched is varied and builds on many types of study designs and methodologies. A little more than half of the included studies did not provide a clear motivation for implementing an AI system, which is a key factor for successful adoption of AI in health care [105]. The lack of a clear motivation indicates poor alignment with well-defined needs from clinical practice and risks reinforcing a technology-focused logic regarding implementation of AI in health care. This observation might reflect the lack of consistent understanding of what is meant by implementation of AI in daily practice and a lack of methodological consistency in how such implementations should be researched and reported. Most of the studies either had a technical or computational understanding of implementation or viewed implementation in terms of the effectiveness of the intervention. There was not much focus on the actual process of implementation studies but more on presenting cases of implementation. This indicates the relatively nascent nature of evidence in this field and is similar to other studies, which highlights that many of the publications on AI in health care focus on the methods and technical aspects of applying the AI model to clinical scenarios but provide very little information on the actual process of its implementation in practice [51,99].

Despite the limited focus in the studies on researching the implementation process, our inductive analysis identified the following implementation elements: cocreation, designing nondisruptive workflows, maintaining a learning focus, communication, contextualization, leadership and conducive organizational structure, trainings, and enforcement or incentivization of AI use. These aspects are not unique to AI but have been highlighted as important interventions for the adoption of all digital technologies, including AI; for example, the involvement of end users in the design and implementation of IT services and applications forms the basis of user-centered design, which is seen as an important driver of uptake of digital technologies [106]. The commitment, involvement, and accountability of leaders is also a well-known factor for successful implementation in practice [107]. Seamless integration with existing workflows was another factor highlighted as central to adoption of AI systems. This finding is consistent with the fact that most studied cases of AI system implementation were based on decision support systems that have no action autonomy and can be conveniently incorporated into routine workflows. However, it is challenging to draw generalized conclusions on the AI implementation strategies from such systems because they introduce incremental improvements in the workflows and do not represent more disruptive types of AI systems; for example, those with high action autonomy.

The findings of this study corroborate the recent work by Gama et al [37] regarding the uncertainty of what should be considered AI and the notion that our understanding of implementation is still in the early stages of development. We would add that this understanding is made even more complex by the lack of agreement on what is meant by the term implementation. We rejected numerous studies during the screening because the term implementation was used in a computational sense; for example, the product concept or requirements were *implemented* as a code, or the coded algorithm was *implemented* using an existing data set. Even in studies involving real-world settings, the term was used to mean execution of a plan without reflection on the process of execution. The focus of implementation as an intentional effort designed to change routine practice, adapt interventions, or increase the uptake of interventions into routine practice was scarce in the published literature.

Limitations and Methodological Considerations

The strengths of this study include the substantial number of records reviewed and the rigor observed during the screening process. The search strategy was comprehensive and broad, and covered 5 different electronic databases. However, we did not include a broader search of the gray literature that would have undoubtedly captured additional cases and potentially identified more cases representing ongoing or completed implementation projects not yet published in the research literature. As we aimed to investigate the experiences from implementation in clinical practice, during screening we removed clinical trials, case reports, pilots, feasibility studies, and other forms of limited and controlled introduction of AI applications in practice. We

expect there to be a lag between the work of technology companies and care providers and subsequent academic publications. However, because of the number of records we identified and the previously found extensive availability of opinion-based articles in the literature in the form of perspectives, insights, and narrative reviews [37], we made a conscious choice to focus on peer-reviewed articles. Although this procedure might risk excluding relevant knowledge from smaller or unsuccessful implementation attempts or other research adjacent to implementation processes, we delimited the results to the literature based on actual experiences from implementation in everyday clinical practice.

Our initial screening of title and abstracts did not require decisions by 2 reviewers, but all decisions in the full-text screening were confirmed in pairs. We deliberately worked to maintain consistency and mitigate individual variation through biweekly meetings where we worked to establish a psychologically safe environment that encouraged all authors to raise or flag doubts, discuss the application of exclusion criteria, or consider differing interpretations. When in doubt, we would backtrack or repeat without blame, and all conflicts and uncertainties were resolved through discussion until consensus was reached. Additional meetings were held with other experts in the domain to ensure methodological rigor. Although the Arksey and O'Malley framework for scoping reviews [39] does not include a quality appraisal, we would recommend that future authors consider doing so as the number of articles that carefully consider implementation increases.

Conclusions

The current body of empirical evidence demonstrates a dissonance between research and practice needs. On the one hand, conceptual and methodological AI research builds on large promises of AI to revolutionize health care and problematizes its slow uptake into practice. On the other hand, the current empirically supported knowledge derives mostly from implementations of AI systems with low action autonomy and highlights lessons on the implementation process that are typical of implementations of other types of information systems. Further research is needed on the more disruptive types of AI systems being implemented in routine care to identify those aspects of implementation unique to AI. This highlights the need for future research to advance in two main streams: (1) to empirically study the implementation processes of various types of AI systems in health care practice and (2) to support empirical research and practical implementations by developing and disseminating an AI-specific implementation framework that would take into account some of the unique aspects related to uptake of AI in health care, such as building trust, addressing transparency issues, developing explainable and interpretable solutions, and addressing ethical concerns around privacy and data protection.

Acknowledgments

The authors would like to thank Per Nilsen and Julie Reed for contributing critical input to the final draft of the manuscript. This work was financially supported by the Knowledge Foundation. The funders had no involvement in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

Authors' Contributions

All authors participated in the design of the study. IL, CS, and JMN retrieved the records from databases. IL, CS, JMN, MS, and MN participated in the screening process. MS, MN, and IL extracted the data. All authors were involved in the analysis and interpretation of data. MS originally drafted the manuscript as a master's thesis, with CS and IL as supervisors and input from all authors. All authors thereafter drafted and revised the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy for the study.
[DOCX File , 29 KB - jmir_v24i10e40238_app1.docx]

Multimedia Appendix 2
Overview of articles included in the scoping review (N=45).
[DOCX File , 21 KB - jmir_v24i10e40238_app2.docx]

Multimedia Appendix 3
Types of artificial intelligence (AI) systems implemented and main focus of research. (A) Types of AI technologies implemented, classified according to the Organisation for Economic Co-operation and Development Framework. (B) Types of tasks performed by AI in health care across the included studies. (C) Level of action autonomy in the AI implemented. (D) Overall focus of the paper and results.
[PNG File , 128 KB - jmir_v24i10e40238_app3.png]

References

1. No longer science fiction, AI and robotics are transforming healthcare. PwC. URL: https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html [accessed 2022-05-30]
2. Das R. Five Technologies That Will Disrupt Healthcare By 2020. Forbes. 2016 Mar 30. URL: https://www.forbes.com/sites/reenitadas/2016/03/30/top-5-technologies-disrupting-healthcare-by-2020/ [accessed 2022-05-30]
3. Bitterman DS, Aerts HJ, Mak RH. Approaching autonomy in medical artificial intelligence. Lancet Digit Health 2020 Sep;2(9):e447-e449 [FREE Full text] [doi: 10.1016/S2589-7500(20)30187-4] [Medline: 33328110]
4. Langlotz CP. Will artificial intelligence replace radiologists? Radiol Artif Intell 2019 May;1(3):e190058 [FREE Full text] [doi: 10.1148/ryai.2019190058] [Medline: 33937794]
5. Shuaib A, Arian H, Shuaib A. The increasing role of artificial intelligence in health care: will robots replace doctors in the future? Int J Gen Med 2020 Oct 19;13:891-896 [FREE Full text] [doi: 10.2147/IJGM.S268093] [Medline: 33116781]
6. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc Manage Forum 2020 Jan;33(1):10-18. [doi: 10.1177/0840470419873123] [Medline: 31550922]
7. 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]
8. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
9. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 2021 Apr 10;21(1):125 [FREE Full text] [doi: 10.1186/s12911-021-01488-9] [Medline: 33836752]
10. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. JAMA Surg 2020 Feb 01;155(2):148-158 [FREE Full text] [doi: 10.1001/jamasurg.2019.4917] [Medline: 31825465]
11. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, Carter R, Chareonthaitawee P, Jouni H, et al. Artificial intelligence in cardiology: present and future. Mayo Clin Proc 2020 May;95(5):1015-1039. [doi: 10.1016/j.mayocp.2020.01.038] [Medline: 32370835]
12. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HT, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. J Med Internet Res 2021 Sep 03;23(9):e29839 [FREE Full text] [doi: 10.2196/29839] [Medline: 34477556]
13. Checcucci E, Autorino R, Cacciamani GE, Amparore D, De Cillis S, Piana A, Uro-technology and SoMe Working Group of the Young Academic Urologists Working Party of the European Association of Urology. Artificial intelligence and neural networks in urology: current clinical applications. Minerva Urol Nefrol 2020 Feb;72(1):49-57 [FREE Full text] [doi: 10.23736/S0393-2249.19.03613-0] [Medline: 31833725]
14. Dutt S, Sivaraman A, Savoy F, Rajalakshmi R. Insights into the growing popularity of artificial intelligence in ophthalmology. Indian J Ophthalmol 2020 Jul;68(7):1339-1346 [FREE Full text] [doi: 10.4103/ijo.IJO_1754_19] [Medline: 32587159]
15. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
16. Chen ZH, Lin L, Wu CF, Li CF, Xu RH, Sun Y. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. Cancer Commun (Lond) 2021 Nov;41(11):1100-1115 [FREE Full text] [doi: 10.1002/cac2.12215] [Medline: 34613667]
17. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan G, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond) 2020 Mar;34(3):451-460 [FREE Full text] [doi: 10.1038/s41433-019-0566-0] [Medline: 31488886]
18. Ozsahin I, Sekeroglu B, Musa MS, Mustapha MT, Uzun Ozsahin D. Review on diagnosis of COVID-19 from chest CT images using artificial intelligence. Comput Math Methods Med 2020 Sep 26;2020:9756518 [FREE Full text] [doi: 10.1155/2020/9756518] [Medline: 33014121]
19. Romm EL, Tsigelny IF. Artificial intelligence in drug treatment. Annu Rev Pharmacol Toxicol 2020 Jan 06;60:353-369. [doi: 10.1146/annurev-pharmtox-010919-023746] [Medline: 31348869]
20. Ho D. Artificial intelligence in cancer therapy. Science 2020 Feb 28;367(6481):982-983. [doi: 10.1126/science.aaz3023] [Medline: 32108102]
21. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May 16;395(10236):1579-1586 [FREE Full text] [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
22. Lorenzi NM, Riley RT, Blyth AJ, Southon G, Dixon BJ. Antecedents of the people and organizational aspects of medical informatics: review of the literature. J Am Med Inform Assoc 1997;4(2):79-93 [FREE Full text] [doi: 10.1136/jamia.1997.0040079] [Medline: 9067874]
23. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003 Jun;36(3):218-227 [FREE Full text] [doi: 10.1016/j.jbi.2003.09.002] [Medline: 14615230]

24. Gooding P, Kariotis T. Ethics and law in research on algorithmic and data-driven technology in mental health care: scoping review. JMIR Ment Health 2021 Jun 10;8(6):e24668 [FREE Full text] [doi: 10.2196/24668] [Medline: 34110297]
25. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. Int J Med Inform 2022 May;161:104738. [doi: 10.1016/j.ijmedinf.2022.104738] [Medline: 35299098]
26. Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp 2019 Dec 10;7(1):70 [FREE Full text] [doi: 10.1186/s40635-019-0286-6] [Medline: 31823128]
27. 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]
28. Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: a systematic review. Soc Sci Med 2022 Mar;296:114782 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114782] [Medline: 35152047]
29. Trocin C, Mikalef P, Papamitsiou Z, Conboy K. Responsible AI for digital health: a synthesis and a research agenda. Inf Syst Front (forthcoming) 2021 Jun 26 [FREE Full text] [doi: 10.1007/s10796-021-10146-4]
30. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
31. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform 2020 Jul 24;8(7):e18599 [FREE Full text] [doi: 10.2196/18599] [Medline: 32706688]
32. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
33. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res 2020 Feb 20;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]
34. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022 Jan 10;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
35. Mirbabaie M, Stieglitz S, Frick NR. Artificial intelligence in disease diagnostics: a critical review and classification on the current state of research guiding future direction. Health Technol 2021 May 10;11(4):693-731 [FREE Full text] [doi: 10.1007/s12553-021-00555-5]
36. Zidaru T, Morrow EM, Stockley R. Ensuring patient and public involvement in the transition to AI-assisted mental health care: a systematic scoping review and agenda for design justice. Health Expect 2021 Aug;24(4):1072-1124 [FREE Full text] [doi: 10.1111/hex.13299] [Medline: 34118185]
37. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215 [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
38. Alhashmi SF, Alshurideh M, Al Kurdi B, Salloum SA. A systematic review of the factors affecting the artificial intelligence implementation in the health care sector. In: Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Vision. 2020 Presented at: AICV '20; April 8–10, 2020; Cairo, Egypt p. 37-49 URL: https://link.springer.com/chapter/10.1007/978-3-030-44289-7_4 [doi: 10.1007/978-3-030-44289-7_4]
39. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
40. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
41. Svedberg P, Reed J, Nilsen P, Barlow J, Macrae C, Nygren J. Toward successful implementation of artificial intelligence in health care practice: protocol for a research program. JMIR Res Protoc 2022 Mar 09;11(3):e34920 [FREE Full text] [doi: 10.2196/34920] [Medline: 35262500]
42. Recommendation of the Council on Artificial Intelligence (OECD Legal Instruments. OECD/ LEGAL/O449). Organisation for Economic Co-operation and Development. 2019 May 22. URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText [accessed 2022-02-26]
43. Ethics and Governance of Artificial Intelligence for Health: WHO guidance. World Health Organization. 2021 Jun 28. URL: http://apps.who.int/bookorders [accessed 2022-02-16]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Ng R, Tan KB. Implementing an individual-centric discharge process across Singapore public hospitals. Int J Environ Res Public Health 2021 Aug 17;18(16):8700 [FREE Full text] [doi: 10.3390/ijerph18168700] [Medline: 34444448]
46. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020 Jul 15;8(7):e15182 [FREE Full text] [doi: 10.2196/15182] [Medline: 32673244]

47. Kashyap S, Morse KE, Patel B, Shah NH. A survey of extant organizational and computational setups for deploying predictive models in health systems. J Am Med Inform Assoc 2021 Oct 12;28(11):2445-2450 [FREE Full text] [doi: 10.1093/jamia/ocab154] [Medline: 34423364]
48. McKillop M, South B, Preininger A, Mason M, Jackson G. Leveraging conversational technology to answer common COVID-19 questions. J Am Med Inform Assoc 2021 Mar 18;28(4):850-855 [FREE Full text] [doi: 10.1093/jamia/ocaa316] [Medline: 33517402]
49. Moorman LP. Principles for real-world implementation of bedside predictive analytics monitoring. Appl Clin Inform 2021 Aug;12(4):888-896 [FREE Full text] [doi: 10.1055/s-0041-1735183] [Medline: 34553360]
50. Baxter SL, Bass JS, Sitapati AM. Barriers to implementing an artificial intelligence model for unplanned readmissions. ACI open 2020 Jul;4(2):e108-e113 [FREE Full text] [doi: 10.1055/s-0040-1716748] [Medline: 33274314]
51. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform 2020 May;137:104072. [doi: 10.1016/j.ijmedinf.2019.104072] [Medline: 32200295]
52. Lai L, Wittbold K, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020 Dec;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]
53. Strohm L, Hehakaya C, Ranschaert ER, Boon WP, Moors EH. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020 Oct;30(10):5525-5532 [FREE Full text] [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
54. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. Implementation of artificial intelligence-based clinical decision support to reduce hospital readmissions at a regional hospital. Appl Clin Inform 2020 Aug;11(4):570-577 [FREE Full text] [doi: 10.1055/s-0040-1715827] [Medline: 32877943]
55. Reis L, Maier C, Mattke J, Creutzenberg M, Weitzel T. Addressing user resistance would have prevented a healthcare AI project failure. MIS Q Exec 2020 Dec;19(4):279-296 [FREE Full text] [doi: 10.17705/2msqe.00038]
56. Sun TQ. Adopting artificial intelligence in public healthcare: the effect of social power and learning algorithms. Int J Environ Res Public Health 2021 Dec 01;18(23):12682 [FREE Full text] [doi: 10.3390/ijerph182312682] [Medline: 34886404]
57. Wijnhoven F. Organizational learning for intelligence amplification adoption: lessons from a clinical decision support system adoption project. Inf Syst Front 2021 Oct 09;24(3):731-744 [FREE Full text] [doi: 10.1007/s10796-021-10206-9]
58. Thurzo A, Kurilová V, Varga I. Artificial intelligence in orthodontic smart application for treatment coaching and its impact on clinical performance of patients monitored with AI-TeleHealth system. Healthcare (Basel) 2021 Dec 07;9(12):1695 [FREE Full text] [doi: 10.3390/healthcare9121695] [Medline: 34946421]
59. Gonçalves LS, Amaro ML, Romero AD, Schamne FK, Fressatto JL, Bezerra CW. Implementation of an artificial intelligence algorithm for sepsis detection. Rev Bras Enferm 2020 Apr 9;73(3):e20180421 [FREE Full text] [doi: 10.1590/0034-7167-2018-0421] [Medline: 32294705]
60. Damoah IS, Ayakwah A, Tingbani I. Artificial intelligence (AI)-enhanced medical drones in the healthcare supply chain (HSC) for sustainability development: a case study. J Clean Prod 2021 Dec;328:129598. [doi: 10.1016/j.jclepro.2021.129598]
61. Lacey G, Zhou J, Li X, Craven C, Gush C. The impact of automatic video auditing with real-time feedback on the quality and quantity of handwash events in a hospital setting. Am J Infect Control 2020 Feb;48(2):162-166 [FREE Full text] [doi: 10.1016/j.ajic.2019.06.015] [Medline: 31358419]
62. Saverino A, Baiardi P, Galata G, Pedemonte G, Vassallo C, Pistarini C. The challenge of reorganizing rehabilitation services at the time of COVID-19 pandemic: a new digital and artificial intelligence platform to support team work in planning and delivering safe and high quality care. Front Neurol 2021 Apr 29;12:643251 [FREE Full text] [doi: 10.3389/fneur.2021.643251] [Medline: 33995247]
63. Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. Radiat Oncol 2021 Jun 08;16(1):101 [FREE Full text] [doi: 10.1186/s13014-021-01831-4] [Medline: 34103062]
64. Herman B, Sirichokchatchawan W, Nantasenamat C, Pongpanich S. Artificial intelligence in overcoming rifampicin resistant-screening challenges in Indonesia: a qualitative study on the user experience of CUHAS-ROBUST. J Health Res (forthcoming) 2021 Jun 15 [FREE Full text] [doi: 10.1108/jhr-11-2020-0535]
65. Chong J, Curtain C, Gad F, Passam F, Soo G, Levy R, et al. Development and implementation of venous thromboembolism stewardship across a hospital network. Int J Med Inform 2021 Nov;155:104575. [doi: 10.1016/j.ijmedinf.2021.104575] [Medline: 34560489]
66. Tan JR, Cheong EH, Chan LP, Tham WP. Implementation of an artificial intelligence-based double read system in capturing pulmonary nodule discrepancy in CT studies. Curr Probl Diagn Radiol 2021;50(2):119-122. [doi: 10.1067/j.cpradiol.2020.07.006] [Medline: 32839067]
67. Petitgand C, Motulsky A, Denis JL, Régis C. Investigating the barriers to physician adoption of an artificial intelligence-based decision support system in emergency care: an interpretative qualitative study. Stud Health Technol Inform 2020 Jun 16;270:1001-1005. [doi: 10.3233/SHTI200312] [Medline: 32570532]

68. Chonde DB, Pourvaziri A, Williams J, McGowan J, Moskos M, Alvarez C, et al. RadTranslate: an artificial intelligence-powered intervention for urgent imaging to enhance care equity for patients with limited English proficiency during the COVID-19 pandemic. J Am Coll Radiol 2021 Jul;18(7):1000-1008 [FREE Full text] [doi: 10.1016/j.jacr.2021.01.013] [Medline: 33609456]
69. García Bermúdez I, González Manso M, Sánchez Sánchez E, Rodríguez Hita A, Rubio Rubio M, Suárez Fernández C. Usefulness and acceptance of telephone monitoring by a virtual assistant for patients with COVID-19 following discharge. Rev Clin Esp (Barc) 2021 Oct;221(8):464-467 [FREE Full text] [doi: 10.1016/j.rceng.2021.01.007] [Medline: 34172430]
70. Tamposis I, Tsougos I, Karatzas A, Vassiou K, Vlychou M, Tzortzis V. PCaGuard: a software platform to support optimal management of prostate cancer. Appl Clin Inform 2022 Jan;13(1):91-99 [FREE Full text] [doi: 10.1055/s-0041-1741481] [Medline: 35045583]
71. Mohamed M, Ferguson DJ, Venugopal A, Alam MK, Makki L, Vaid NR. An artificial intelligence based referral application to optimize orthodontic referrals in a public oral healthcare system. Semin Orthod 2021 Jun 1;27(2):157-163. [doi: 10.1053/j.sodo.2021.05.011]
72. Morales HM, Guedes M, Silva JS, Massuda A. COVID-19 in Brazil-preliminary analysis of response supported by artificial intelligence in municipalities. Front Digit Health 2021 Jun 17;3:648585 [FREE Full text] [doi: 10.3389/fdgth.2021.648585] [Medline: 34713121]
73. Davis MA, Rao B, Cedeno PA, Saha A, Zohrabian VM. Machine learning and improved quality metrics in acute intracranial hemorrhage by noncontrast computed tomography. Curr Probl Diagn Radiol 2022;51(4):556-561. [doi: 10.1067/j.cpradiol.2020.10.007] [Medline: 33243455]
74. Snowdon JL, Robinson B, Staats C, Wolsey K, Sands-Lincoln M, Strasheim T, et al. Empowering caseworkers to better serve the most vulnerable with a cloud-based care management solution. Appl Clin Inform 2020 Aug;11(4):617-621 [FREE Full text] [doi: 10.1055/s-0040-1715894] [Medline: 32969000]
75. Schlicher J, Metsker M, Shah H, Demirkan H. From NASA to healthcare: real-time data analytics (mission control) is reshaping healthcare services. Perspect Health Inf Manag 2021 Oct 1;18(4):1g [FREE Full text] [Medline: 34975356]
76. O'Neill TJ, Xi Y, Stehel E, Browning T, Ng YS, Baker C, et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. Radiol Artif Intell 2020 Nov 18;3(2):e200024 [FREE Full text] [doi: 10.1148/ryai.2020200024] [Medline: 33937858]
77. Bennett CC. Clinical productivity system - a decision support model. Int J Prod Performance Manag 2011;60(3):311-319 [FREE Full text] [doi: 10.1108/17410401111112014]
78. Schuh C, de Bruin JS, Seeling W. Clinical decision support systems at the Vienna general hospital using Arden syntax: design, implementation, and integration. Artif Intell Med 2018 Nov;92:24-33. [doi: 10.1016/j.artmed.2015.11.002] [Medline: 26706047]
79. Anand V, Carroll AE, Biondich PG, Dugan TM, Downs SM. Pediatric decision support using adapted Arden syntax. Artif Intell Med 2018 Nov;92:15-23 [FREE Full text] [doi: 10.1016/j.artmed.2015.09.006] [Medline: 26547523]
80. Wen A, Fu S, Moon S, El Wazir M, Rosenbaum A, Kaggal VC, et al. Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. NPJ Digit Med 2019 Dec 17;2:130 [FREE Full text] [doi: 10.1038/s41746-019-0208-8] [Medline: 31872069]
81. Rath S, Rajaram K, Mahajan A. Integrated anesthesiologist and room scheduling for surgeries: methodology and application. Operation Res 2017 Jul 20;65(6):1460-1478 [FREE Full text] [doi: 10.1287/opre.2017.1634]
82. Rais A, Alvelos F, Figueiredo J, Nobre A. Optimization of logistics services in hospitals. Intl Trans Op Res 2018 Jan;25(1):111-132 [FREE Full text] [doi: 10.1111/itor.12370]
83. Dios M, Molina-Pariente JM, Fernandez-Viagas V, Andrade-Pineda JL, Framinan JM. A decision support system for operating room scheduling. Comput Ind Eng 2015 Oct;88(C):430-443 [FREE Full text] [doi: 10.1016/j.cie.2015.08.001]
84. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care. Int J Med Inform 2012 Aug;81(8):521-526. [doi: 10.1016/j.ijmedinf.2012.03.002] [Medline: 22483528]
85. Cruz NP, Canales L, Muñoz JG, Pérez B, Arnott I. Improving adherence to clinical pathways through natural language processing on electronic medical records. Stud Health Technol Inform 2019 Aug 21;264:561-565. [doi: 10.3233/SHTI190285] [Medline: 31437986]
86. Semenov I, Kopanitsa G. Development of a clinical decision support system for the patients of a laboratory service. Stud Health Technol Inform 2016;228:90-94. [Medline: 27577348]
87. Campion Jr TR, Waitman LR, Lorenzi NM, May AK, Gadd CS. Barriers and facilitators to the use of computer-based intensive insulin therapy. Int J Med Inform 2011 Dec;80(12):863-871 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.10.003] [Medline: 22019280]
88. Kalil AJ, Dias VM, Rocha CD, Morales HM, Fressatto JL, Faria RA. Sepsis risk assessment: a retrospective analysis after a cognitive risk management robot (Robot Laura®) implementation in a clinical-surgical unit. Res Biomed Eng 2018 Nov 22;34(4):310-316 [FREE Full text] [doi: 10.1590/2446-4740.180021]
89. Sukums F, Mensah N, Mpembeni R, Massawe S, Duysburgh E, Williams A, et al. Promising adoption of an electronic clinical decision support system for antenatal and intrapartum care in rural primary healthcare facilities in sub-Saharan

Africa: the QUALMAT experience. Int J Med Inform 2015 Sep;84(9):647-657. [doi: 10.1016/j.ijmedinf.2015.05.002] [Medline: 26073076]
90. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
91. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
92. Nonaka I. A dynamic theory of organizational knowledge creation. Organ Sci 1994 Feb;5(1):14-37 [FREE Full text] [doi: 10.1287/orsc.5.1.14]
93. Hadley TD, Pettit RW, Malik T, Khoei AA, Salihu HM. Artificial intelligence in global health -a framework and strategy for adoption and sustainability. Int J MCH AIDS 2020;9(1):121-127 [FREE Full text] [doi: 10.21106/ijma.296] [Medline: 32123635]
94. Arnold MH. Teasing out artificial intelligence in medicine: an ethical critique of artificial intelligence and machine learning in medicine. J Bioeth Inq 2021 Mar;18(1):121-139 [FREE Full text] [doi: 10.1007/s11673-020-10080-1] [Medline: 33415596]
95. 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]
96. Higgins D, Madai VI. From bit to bedside: a practical framework for artificial intelligence product development in healthcare. Adv Intell Syst 2020 Oct;2(10):2000052 [FREE Full text] [doi: 10.1002/aisy.202000052]
97. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018 Aug 29;3(4):e000798 [FREE Full text] [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
98. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020 May;26(5):584-595 [FREE Full text] [doi: 10.1016/j.cmi.2019.09.009] [Medline: 31539636]
99. Kueper JK, Terry AL, Zwarenstein M, Lizotte DJ. Artificial intelligence and primary care research: a scoping review. Ann Fam Med 2020 May;18(3):250-258 [FREE Full text] [doi: 10.1370/afm.2518] [Medline: 32393561]
100. Rahman MM, Khatun F, Uzzaman A, Sami SI, Bhuiyan MA, Kiong TS. A comprehensive study of artificial intelligence and machine learning approaches in confronting the coronavirus (COVID-19) pandemic. Int J Health Serv 2021 Oct;51(4):446-461 [FREE Full text] [doi: 10.1177/00207314211017469] [Medline: 33999732]
101. Quinn TP, Jacobs S, Senadeera M, Le V, Coghlan S. The three ghosts of medical AI: can the black-box present deliver? Artif Intell Med 2022 Feb;124:102158. [doi: 10.1016/j.artmed.2021.102158] [Medline: 34511267]
102. Kort om Artificiell intelligens i hälso- och sjukvården. Statens medicinsk-etiska råd. 2020. URL: https://smer.se/2020/03/31/kort-om-artificiell-intelligens-i-halso-och-sjukvarden/ [accessed 2022-05-30]
103. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
104. Montani S, Striani M. Artificial intelligence in clinical decision support: a focused literature survey. Yearb Med Inform 2019 Aug;28(1):120-127 [FREE Full text] [doi: 10.1055/s-0039-1677911] [Medline: 31419824]
105. Saria S. Five keys for using artificial intelligence in care delivery, and five common mistakes made in evaluating AI tools. Not All AI Is Created Equal: Strategies for Safe and Effective Adoption. NEJM Catalyst. 2022 Mar 23. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.22.0075 [accessed 2022-05-30]
106. Bernstein ML, McCreless T, Côté MJ. Five constants of information technology adoption in healthcare. Hosp Top 2007;85(1):17-25. [doi: 10.3200/HTPS.85.1.17-26] [Medline: 17405421]
107. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

Abbreviations

AI: artificial intelligence
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by G Eysenbach; submitted 12.06.22; peer-reviewed by M Gordon, , R Damaševičius; comments to author 01.07.22; revised version received 19.08.22; accepted 30.08.22; published 05.10.22.

Please cite as:
Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM
Artificial Intelligence Applications in Health Care Practice: Scoping Review
J Med Internet Res 2022;24(10):e40238
URL: https://www.jmir.org/2022/10/e40238
doi:10.2196/40238
PMID:36197712

Review

Cultural and Digital Health Literacy Appropriateness of App- and Web-Based Systems Designed for Pregnant Women With Gestational Diabetes Mellitus: Scoping Review

Yosefa Birati[1], PhD; Enav Yefet[1,2], MD, PhD; Yuri Perlitz[1,2], MD; Naim Shehadeh[3], MD; Sivan Spitzer[1], PhD

[1]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel

[2]Department of Obstetrics and Gynecology, Baruch Padeh Medical Center Poriya, Tiberias, Israel

[3]Institute of Endocrinology, Diabetes and Metabolism, Rambam Health Care Campus, Haifa, Israel

Corresponding Author:
Yosefa Birati, PhD
The Azrieli Faculty of Medicine
Bar-Ilan University
8 Henrietta Szold Street
Safed, 1311502
Israel
Phone: 972 522684065
Email: ebirati1@hotmail.com

Abstract

Background: The prevalence of women diagnosed with gestational diabetes mellitus (GDM) is increasing dramatically. Mobile technologies to enhance patient self-management offer many advantages for women diagnosed with GDM. However, to our knowledge, although mobile health (mHealth) and telemedicine systems for GDM management exist, evidence on their cultural and digital health literacy appropriateness levels is limited.

Objective: This review aimed to search and assess the literature on mHealth and telemedicine systems designed for women diagnosed with GDM. Our assessment of these technologies focused on their cultural and digital health literacy appropriateness as well as the systems' effectiveness in improving glycemic control and maternal and infant outcomes.

Methods: We conducted a scoping review using a framework adapted from Arksey and O'Malley. Four electronic databases were searched for relevant studies: PubMed, MEDLINE (EBSCO), Web of Science, and Scopus. The databases were searched between January 2010 and January 2022. The inclusion criteria were pregnant women diagnosed with GDM, use of telemedicine for monitoring and management, and vulnerable or disadvantaged patients. We used terms related to mobile apps and telemedicine: *GDM*, *vulnerable populations*, *periphery*, *cultural appropriateness*, and *digital health literacy*. Studies were screened and selected independently by 2 authors. We extracted the study data on a Microsoft Excel charting table and categorized them into final themes. The results were categorized according to the cultural and digital health literacy features presented.

Results: We identified 17 studies that reported on 12 telemedicine and mHealth app interventions. We assessed the studies in three domains: cultural appropriateness, digital health literacy, and maternal and infant outcomes. In the literature, we found that existing digital technologies may improve glycemic control and diabetes self-management. However, there is a lack of assessment of cultural and digital health literacy appropriateness for pregnant women diagnosed with GDM. Considerations in app design regarding cultural appropriateness were found in only 12% (2/17) of the studies, and only 25% (3/12) of the interventions scored ≥3 out of 5 in our assessment of digital health literacy.

Conclusions: mHealth and telemedicine can be an effective platform to improve the clinical management of women with GDM. Although studies published on the use of mHealth and telemedicine systems exist, there is a limited body of knowledge on the digital health literacy and cultural appropriateness of the systems designed for women diagnosed with GDM. In addition, as our study was restricted to the English language, relevant studies may have been excluded. Further research is needed to evaluate, design, and implement better tailored apps regarding cultural and digital literacy appropriateness for enhancing pregnant women's self-management as well as the effectiveness of these apps in improving maternal and infant health outcomes.

(J Med Internet Res 2022;24(10):e37844) doi:10.2196/37844

KEYWORDS

gestational diabetes mellitus; maternal health; mobile health; mHealth; mobile apps; mobile phone; telemedicine; culture; health literacy; vulnerable populations; pregnancy outcome

Introduction

Background

The prevalence of women diagnosed with diabetes during pregnancy has substantially increased over the last decade. The International Diabetes Federation 2019 report estimated that 20 million women developed hyperglycemia during pregnancy, 84% of them because of gestational diabetes mellitus (GDM) [1]. GDM is associated with significant risk and can lead to grave adverse perinatal outcomes and long-term health complications for both mothers and their offspring [2-4]. Self-care and changes in lifestyle are essential for adequate glycemic control and prevention of unfavorable maternal-infant health outcomes [5,6]. GDM management requires women to implement medical nutrition therapy, self-monitor their blood glucose levels, manage weight gain, and perform physical exercise [7-9]. These self-management tasks are complex and pose a significant self-care burden for pregnant women, especially those who are diagnosed for the first time.

Digital technology solutions have been introduced to support and improve women's management and outcomes while decreasing the need for direct physician-patient contact. Digital health platforms include mobile health (mHealth) apps, telehealth, and telemedicine, and the information can be delivered through a wide range of technologies such as web-based services, mobile devices, and software systems. These platforms are perceived as a way to reduce disparities in access and quality of care for patients living in rural areas [10]. However, although these technologies have many potential advantages, the extent to which they address the needs of women with diverse communication competencies, culture, and language and different health and digital literacy levels remains unclear [11,12].

Digital health literacy is defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem" [13]. As both health care systems and providers gradually increase their use of health technologies, patients are asked in turn to engage with advanced digitalization, posing additional barriers. Recent studies have shown that low health literacy is positively correlated with deficiencies in diabetes knowledge and self-management among patients with diabetes [14-16] and an increase in health care provider workload [17]. In 2013, the Institute of Medicine published a discussion paper suggesting strategies for improving health literacy and usability by developing health literate apps [18]. However, evidence on the use of mHealth apps and telemedicine systems is based mainly on studies conducted on the general population. Thus, it is not clear if and how levels of digital and health literacy were considered in the development of these systems [19]. Moreover, the Academy of Nutrition and Dietetics guidelines for GDM (2018) suggested that sociocultural assessments such as religious dietary restrictions, food insecurity, or fasting related to religious beliefs should be addressed according to the patient's needs because of their influence on pregnant women's lifestyle and self-management [20]. Cultural appropriateness and cultural sensitivity assessments are essential in the design of digital mHealth apps and telemedicine systems developed for improved GDM management.

Objectives

To our knowledge, reviews assessing digital health literacy and cultural appropriateness of mHealth and telemedicine systems developed for women with GDM have not been conducted. Given the limited evidence, the main objective of this review was to search the literature on mHealth and telemedicine systems designed for women with GDM and assess their cultural and digital health literacy appropriateness as well as the systems' effectiveness in improving glycemic control and maternal and infant outcomes.

Methods

Overview

We conducted this scoping review following the methodological framework proposed by Arksey and O'Malley [21] and Levac et al [22] considering the further refinements made by the Joanna Briggs Institute Reviewers' Manual [23]. The framework by Arksey and O'Malley is based on six essential stages: (1) identifying the research question; (2) searching and identifying relevant studies; (3) selecting the relevant studies; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consulting with stakeholders (optional). We selected the scoping review methodology as our aim was to explore the current body of knowledge regarding GDM mHealth apps tailored for cultural and digital health literacy appropriateness, identify existing knowledge and implementation gaps, and suggest future research needed. Furthermore, the reporting of this scoping review was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for Scoping Reviews (PRISMA-ScR) checklist [24] (Multimedia Appendix 1).

Search Strategy

Four electronic databases—PubMed, MEDLINE (EBSCO), Web of Science, and Scopus—were searched using the following terms: (1) *telemedicine*, (2) *gestational diabetes mellitus*, (3) *target vulnerable populations*, (4) *remote/periphery areas*, (5) *culture appropriate*, and (6) *digital health literacy*. An example of the search strategy and keyword combination for the PubMed database can be found in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

Studies were included based on the inclusion and exclusion criteria and if they met the participants or population, concept, and context mnemonic categorization recommended by the Joanna Briggs Institute for scoping reviews (Table 1).

In our initial search, we examined studies published between January 2010 and October 2021. We updated our search to ensure that any new relevant publications between October 2021 and January 2022 were included. Protocols and feasibility studies for mHealth and web-based system design and development were not included, but they informed our search for publications that presented implementation and study outcomes. Similarly, relevant review publications were not included, but their reference lists were hand searched for additional original papers potentially eligible for inclusion in this scoping review. In addition, we scanned the reference lists of all studies selected for inclusion for additional relevant studies.

Table 1. Inclusion and exclusion criteria.

Categorization	Inclusion criteria	Exclusion criteria
Participants or population	• Pregnant women who were diagnosed with GDM[a]	• Nonpregnant patients • Pregnant women who were not diagnosed with GDM
Concept	• Use of mHealth[b] for GDM monitoring and management • mHealth was considered as telemedicine, mobile phone apps, smartphone apps, and web-based systems	• Use of mHealth telemonitoring for patients who were not diagnosed with GDM (eg, postpartum follow-up, pregnant patients who were diagnosed with type 1 or type 2 diabetes, patients with diabetes who were not pregnant, use of mHealth for pregnant patients following HTN[c], and fetal monitoring)
Context	• Vulnerable or disadvantaged patients or groups (ethnic minorities, migrants, underserved populations, and digital health literacy) • Rural and underserved areas and periphery	• N/A[d]
Type of studies	• Qualitative, quantitative, or mixed methods studies • Observational and experimental, cross-sectional, or longitudinal; RCT[e], nonrandomized, or noncontrolled trials; and case series or case reports	• Conference abstracts, editorials, commentaries, letters to editor, essays, book chapters, and books
Language	• English	• Languages other than English

[a]GDM: gestational diabetes mellitus.

[b]mHealth: mobile health.

[c]HTN: hypertension.

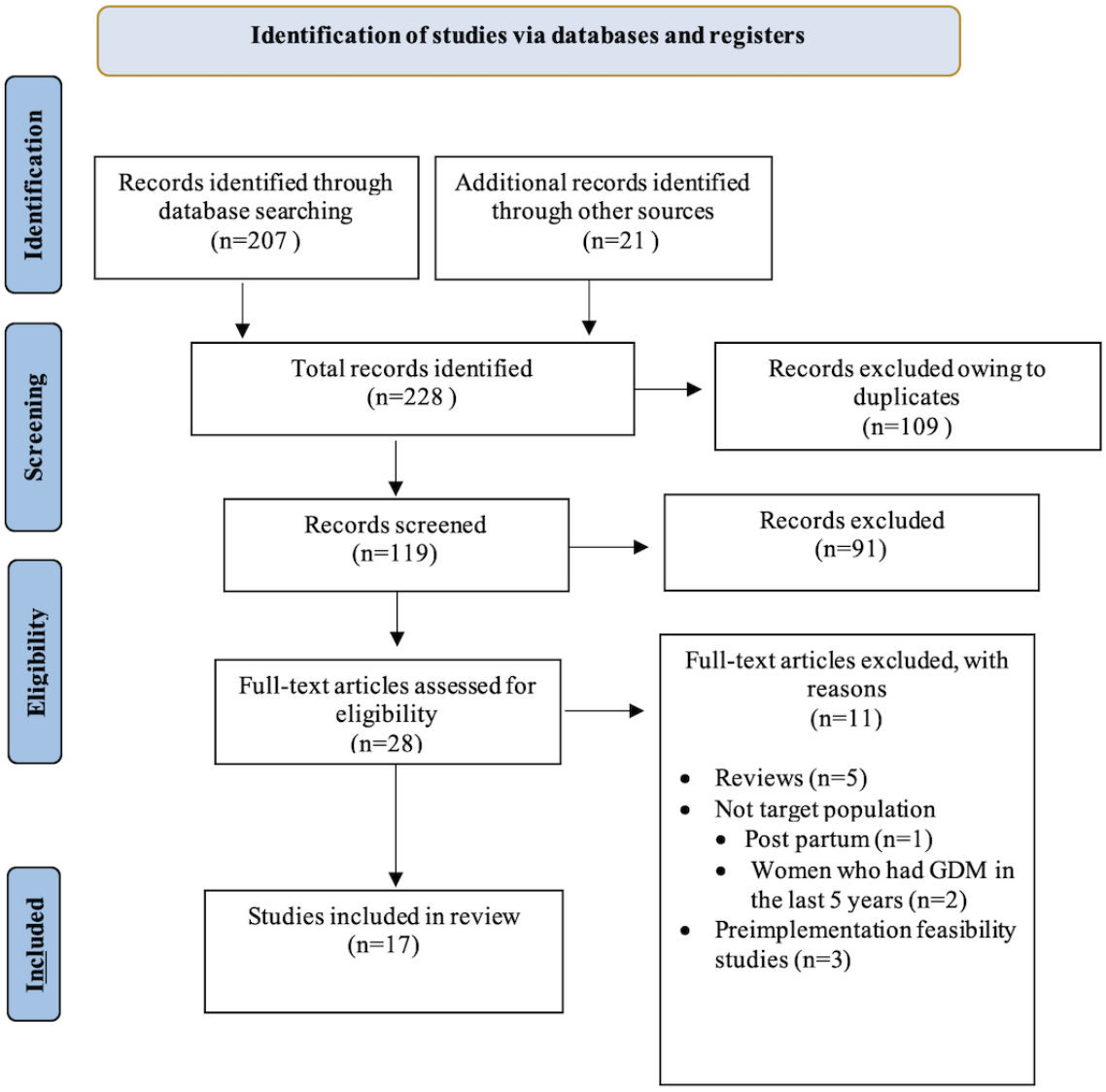
[d]N/A: not applicable.

[e]RCT: randomized controlled trial.

Screening and Selection of Studies

Our initial search of 4 databases yielded 207 results. Our hand search identified 21 additional records. After duplicates were removed, 52.2% (119/228) of publications were reviewed. The selection procedure is presented in the PRISMA flow diagram (Figure 1). The titles and abstracts of 52.2% (119/228) of the articles were screened independently by 2 authors (YB and SS). Following initial screening, of the 119 articles, 91 (76.5%) were excluded, including 5 (4.2%) reviews whose reference lists were searched. The remaining 28 publications' full texts were reviewed and screened for eligibility. A total of 11 publications were excluded (reviews: n=5, 45%; not the target population and women who had GDM but the study was conducted during their postpartum period: n=1, 9%; nonpregnant women who were diagnosed with GDM in the last 5 years: n=2, 18%; and preimplementation usability and feasibility studies: n=3, 27%). Disagreements in the decisions were resolved through discussion and consensus.

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



Charting the Data

The information we derived from the studies included in this scoping review was recorded using a Microsoft Excel (Microsoft Corp) data charting table. The table included general information on the study characteristics (title, year of publication, country, research aims, study design, and population), description of telemonitoring intervention versus usual standard of care, evaluation of digital health literacy, cultural features, outcome measures, and results.

Collating, Summarizing, and Reporting the Results

Selected papers were evaluated thoroughly by both reviewers to identify similarities and differences in mHealth interventions and were summarized for telemedicine and app development and culturally appropriate design. We also evaluated the degree to which these apps addressed health literacy by assessing five digital health literacy features: (1) patients' ability to use a smartphone for chatting, reading, and writing; (2) patient training and guidance on how to use the technology; (3) plain language; (4) display and organization of information (simplified navigation); and (5) the intervention being tested for feasibility and usability. An additional search was conducted to better assess the feasibility and usability of the apps of the interventions included in this scoping review. We searched for previous studies conducted by the authors to evaluate their interventions' feasibility before the final implementation. The outcome measures and significant results were summarized. The data were categorized and organized into final themes.

Results

A total of 17 articles were included for data extraction in this scoping review, as can be seen in the PRISMA flowchart (Figure 1).

Characteristics of the Studies

The included studies were published between 2010 and 2021. Of the 17 articles, 4 (24%) were published in Norway [25-28], and 3 (18%) were published in Spain [29-31], followed by China (n=3, 18%) [32-34], Singapore (n=2, 12%) [35,36], the United States (n=1, 6%) [37], Australia (n=1, 6%) [38], France (n=1,

6%) [39], the United Kingdom (n=1, 6%) [40], and Israel (n=1, 6%) [41]. A total of 35% (6/17) of the studies were multicenter randomized controlled trials (RCTs) [25,26,32,33,37,38], 29% (5/17) were single-center RCTs [29,34,36,40,41], 6% (1/17) were non-RCTs [30], 6% (1/17) had an experimental design [31], 18% (3/17) were qualitative studies [27,28,39], and 6% (1/17) had a mixed methods design [35]. The study interventions included a web-based telemedicine system [30-33,37-39], web-based applications [29,35,36,41], and mHealth apps [25-28,34,40] (Table 2).

Table 2. Characteristics and cultural appropriateness of the 12 apps and systems.

Study	Study design	Country	Study population	Race and ethnicity	Intervention	Cultural appropriateness
Tian et al [32], 2021	RCT[a]	China	• GDM[b] • IG[c]: n=133 • CG[d]: n=136	• Ethnic Han: 96.3% • Other: 3.7%	• WeChat	• +[e] • Individualized guidance for self-management
Yang et al [33], 2018	RCT	China	• GDM • IG: n=57 • CG: n=50 • Normal glucose tolerance: n=50	• N/A[f]	• WeChat	• +/–[g] • Individualized dietary advice
Homko et al [37], 2012	RCT	United States	• GDM • IG: n=36 • CG: n=38	• African American: 30% • White: 37.5% • Latino or Hispanic: 20% • Asian and other: 12.5%	• Web-based system	• –[h]
Garnweidner-Holme et al [26], 2020	RCT	Norway	• GDM • IG: n=95 • CG: n=98	• Norway: 46.6% • Western Europe+United States: 6.7% • Eastern Europe: 9.3% • Asia: 23.3% • Africa: 11.4% • South America: 2.6%	• Pregnant+ app	• + • Norwegian, Urdu, or Somali language and food culture
Borgen et al [25], 2019	RCT	Norway	• GDM • IG: n=115 • CG: n=123	• Norway: 46.8% • Western Europe+United States: 5.9% • Eastern Europe: 8.9% • Asia: 23.6% • Africa: 12.7% • South America: 2.1%	• Pregnant+ app	• + • Norwegian, Urdu, or Somali language
Albert et al [31], 2020	Descriptive—clinical trial	Spain	• GDM • N=20	• N/A	• Web-based system • SineDie	• −
Pérez-Ferre et al [29], 2010	RCT	Spain	• GDM • IG: n=48 • CG: n=49	• IG: • White: 51% • Hispanic: 30.6% • Asian: 6.1% • North African: 4.1% • Other: 8.2% • CG: • White: 56.2% • Hispanic: 37.5% • Asian: 4.2% • North African: 2.1%	• Web internet-based application	• −

Study	Study design	Country	Study population	Race and ethnicity	Intervention	Cultural appropriateness
Surendran et al [35], 2021	RCT and qualitative; mixed methods	Singapore	• GDM • Quantitative data: n=170 • Semistructured interviews: n=14	• Quantitative data • Chinese: 44% • Non-Chinese: 56% • Interviews • Chinese: 57% • Non-Chinese: 43%	• Web-based application • Habits-GDM	• – • Limited common local and ethnic food database (64%) • Food item's name was not worded in the commonly known way, and the imperial measurement (cup) was not familiar to the participants
Carral et al [30], 2015	Prospective interventional study	Spain	• Pregnant women with diabetes • GDM: n=77 • Type 1 DM[i]: n=16 • Type 2 DM: n=11 • IG: n=40 • CG: n=64	• White: 96.2% • Hispanic: 1.9% • North African: 1.9%	• Web-based telemedicine platform • DiabeTIC	• –
Guo et al [34], 2019	RCT	China	• GDM • IG: n=64 • CG: n=60	• N/A	• mHealth[j] app • Dnurse app	• –/+ • Personalized dietary guidance
Mackillop et al [40], 2018	RCT	United Kingdom	• GDM • IG: n=101 • CG: n=102	• IG: • White: 77% • South Asian: 10% • African or Caribbean: 6% • East Asian: 3% • Other: 4% • CG: • White: 78,4% • South Asian: 12.7% • African or Caribbean: 3.9% • East Asian: 1% • Other: 3.9%	• mHealth app	• –
Miremberg et al [41], 2018	RCT	Israel	• GDM • IG: n=64 • CG: n=60	• N/A	• Web-based application	• –
Yew et al [36], 2021	RCT	Singapore	• GDM • IG: n=170 • CG: n=170	• IG: • Chinese: 44.1% • Non-Chinese: 55.9% • CG: • Chinese: 43.5% • Non-Chinese: 56.5%	• Web-based application • Habits-GDM	• –/+ • A database of common foods in Singapore was incorporated into the app • A manuscript written by Surendran et al [35] on the same study found that the food items' names were not worded in the commonly known way, and the imperial measurement (cup) was not familiar to the participants
Skar et al [28], 2019	Qualitative	Norway	• GDM • N=17		• N/A	

Study	Study design	Country	Study population	Race and ethnicity	Intervention	Cultural appropriateness
				• Norway: 59% • Immigrants (Poland, Bulgaria, Turkey, Pakistan, Palestine, and Sweden): 41%		• + • Information about health and nutrition in Norwegian, Urdu, and Somali
Rasekaba et al [38], 2018	RCT	Australia	• GDM • IG: n=61 • CG: n=34	• N/A	• Web-based telemedicine platform • TeleGDM	• −
Khalil [39], 2019	Qualitative	France	• GDM n=5 • Health care providers: • Diabetes specialists (n=8) • Educational nurses (n=8) • Dietitians (n=2) • Gynecologists (n=1) • Midwives (n=1)	• N/A	• Telemonitoring system • myDiabby	• −

[a]RCT: randomized controlled trial.

[b]GDM: gestational diabetes mellitus.

[c]IG: intervention group.

[d]CG: control group.

[e]Considered the issue.

[f]N/A: not applicable.

[g]Partially considered the issue.

[h]Did not consider the issue.

[i]DM: diabetes mellitus.

[j]mHealth: mobile health.

Culturally Appropriate Intervention Design

Cultural appropriateness is defined as "the ability to recognize, understand, and react appropriately to beliefs, values, norms, and behaviors of persons who belong to a cultural or ethnic group that differs from one's own" [42]. We operationally defined culturally appropriate design as the assessment and awareness of the researchers in the intervention's design phase in adapting the app's content and instructions according to the patients' culture, language, religion, customs, and beliefs. Assessment and design of cultural appropriateness was considered in only 17% (2/12) of the telemedicine and app interventions we identified. In the Pregnant+ app study, for example, the authors acknowledged the high prevalence of GDM among immigrant women in Norway and the importance of designing and incorporating linguistic and culturally adapted information. The app was translated into 3 languages and also included preferred food items according to culture [25-28,43]. In the Habits-GDM application design, researchers included a database of common foods in Singapore [36]. However, in a qualitative study that examined the Habits-GDM application users' perceptions, 9 out of the 12 interviewed women stated that the database had limited ethnic foods, and 12 out of 14 women claimed that the measurement units were not familiar to them [35]. Another 12% (2/17) of the studies assumed and stated in their study limitations that their targeted users were users with high levels of cultural literacy, but their sample did not represent high-risk or low socioeconomic groups and, therefore, no cultural modifications were added [34,41]. Another 6% (1/17) of the studies acknowledged the existing gaps among rural and disadvantaged populations, which caused the women to avoid using the telemedicine system [30].

Digital and Health Literacy

Overview

We evaluated the 12 intervention studies according to the 5 digital health literacy features, as presented in Table 3, providing an overall maximum score of 5 on the extent to which the apps addressed the 5 features of digital health literacy (Table 3).

Table 3. Summary of the included studies' digital and health literacy features (N=14).

Study	Ability to use a smartphone for chatting, reading, and writing	Proper training and guidance	Content—plain language	Health content—displayed and organized (simplified navigation)	Test usability	Overall digital and health literacy score out of 5
Tian et al [32], 2021	• –[a] • Inclusion criteria: ability to use a smartphone for chatting, reading, and writing basic Chinese	• N/A[b]	• N/A	• N/A	• –/+[c] • Brief interviews	• 0.5
Yang et al [33], 2018	• – • Study exclusion criteria: inability to operate a mobile phone or WeChat	• +[d] • The research team taught the patients how to use the app	• N/A	• N/A	• N/A	• 1
Homko et al [37], 2012	• – • A total of 7 patients (22%) in the IG[e] never accessed the system	• + • IG received training following installation. A total of 3 patients (20%) needed additional training or to correct technical problems	• N/A	• +	• + • Test usability [44]	• 3
Garnweidner-Holme et al [26], 2020	• +	• – • Relied on the women's own capability to download and use the app	• + • Information was in line with the varying levels of literacy • The content writing, literacy, and visual communication were assessed against the Kreuter message checklist [45]	• + • A multidisciplinary research team and experts in software were involved in the design and development and data privacy and security, as well as a graphic designer and language editor • The content was ordered using 4 icons	• + • A total of 21 pregnant women were involved in the development phase and gave feedback, and adjustments were made [43]	• 4
Borgen et al [25], 2019	• +	• –	• +	• +	• +	• 4
Albert et al [31], 2020	• N/A	• N/A	• N/A	• + • Monitoring data were presented in an electronic logbook	• + • Before insulation, the system was evaluated for validity, safety, and effectiveness [46]	• 2

Study	Ability to use a smartphone for chatting, reading, and writing	Proper training and guidance	Content—plain language	Health content—displayed and organized (simplified navigation)	Test usability	Overall digital and health literacy score out of 5
Pérez-Ferre et al [29], 2010	• − • A total of 10 women were excluded because of inability to understand or comply with the protocol	• − • A total of 5 patients were not able to transmit any data • No further intervention was described to enhance the training • These patients had a lower educational level or difficulties with the language or were not used to new technologies	• N/A	• N/A	• + • Feasibility test [47]	• 1
Surendran et al [35], 2021	• + • Half of the app users (84/170) accessed at least one educational lesson	• N/A	• + • Content of the educational lessons was easy to understand (6/9 women)	• − • All information in one place (3/9 women); difficult to search features in the diet-tracking function (3/12 women)	• −/+ • A qualitative study of the patients' experience was conducted after the study trial [36] • No pilot study was conducted	• 2.5
Carral et al [30], 2015	• + • Satisfaction survey—the platform is easy to use: mean 8.1 (SD 1.5, range 5-10)	• N/A	• N/A	• + Satisfaction survey—navigation through the platform is intuitive: mean 6.7 (SD 3.0, range 1-10); allows me to adequately visualize the information: mean 8.2 (SD 1.8, range 4-10)	• + • A pilot study examined patient satisfaction [48]	• 3
Guo et al [34], 2019	• − • Inclusion criteria: patients with smartphones and proficiency in the use of mobile apps • Most of the patients already had a high level of digital literacy	• N/A	• N/A	• N/A	• N/A	• 0
Mackillop et al [40], 2018	• − • Researchers assumed that the women enrolled in the study had high rates of literacy	• N/A	• N/A		• + • Testing the app development [49]	• 2

Study	Ability to use a smartphone for chatting, reading, and writing	Proper training and guidance	Content—plain language	Health content—displayed and organized (simplified navigation)	Test usability	Overall digital and health literacy score out of 5
				• + • Following the app development test, displays were changed to show BG[f] readings in both graphical and tabular formats with color-coded thresholds [49] • Illustrations were added to on-screen buttons		
Miremberg et al [41], 2018	• − • Inclusion criteria: ability to speak English at least to a level that enabled the women to use the application and communicate with the clinic team (the study was conducted in Israel)	• N/A	• −	• −/+ • Patients reported "high" or "very high" satisfaction with their application-based prenatal care • In total, 80% of the patients reported no difficulty using the application (20% of the patients reported slight difficulty mainly related to the English language barrier)	• N/A	• 0.5
Yew et al [36], 2021	• − • Inclusion criteria: proficiency in English (the study was conducted in Singapore)	• N/A	• N/A	• − • Patients were required to be able to navigate an application	• N/A	• 0
Rasekaba et al [38], 2018	• + • Barriers: financial disadvantage in accessing the service and level of technological literacy [50]	• N/A	• N/A	• N/A	• + • Patient satisfaction survey and provider usability [51]	• 2

[a]Did not consider the issue.

[b]N/A: not applicable.

[c]Partially considered the issue.

[d]Considered the issue.

[e]IG: intervention group.

[f]BG: blood glucose.

Women's Ability to Use a Smartphone for Chatting, Reading, and Writing

We evaluated whether the interventions assessed their patients' levels of digital health literacy before the design and development of the intervention, such as women's ability to use a smartphone for chatting, reading, and writing. The Pregnant+ app was the only one of the 12 interventions in which the content was checked during its app design and development phase using the Kreuter health message checklist with regard to patients' ability to read the app information [45] and that also assessed the suitability of the materials [52]. Following this evaluation, researchers added explanations of diabetes medical terms to the app, and the women received information in accordance with their different literacy levels [43]. In DiabeTIC, a pilot study was conducted on a sample of patients with diabetes not exclusive to GDM to evaluate participants' satisfaction with the telemedicine platform monitoring and metabolic control. A total of 7 out of 32 participants were women diagnosed with GDM. The overall mean score from study participants on the item "the platform is easy to use" was 8.1 (SD 1.5, range 5-10) [30,48]. Rasekaba et al [50] identified in their study that barriers to using the app were due to the women's level of digital literacy and technology proficiency. However, this assessment was conducted after the start of the project [50]. In another 24% (4/17) of the studies, those who were not able to use a mobile phone or had language difficulties were excluded in the enrollment phase [32-34,36,41]. Of these 4 studies, 2 (50%) required proficiency in English even though, in 50% (1/2) of those studies, the country's official language was not English and, in the other 50% (1/2), English was one of 4 official languages [36,41]. Mackillop et al [40,53] assumed that the women they enrolled in their study already had high rates of literacy and did not mention any assessment regarding digital literacy (Table 3).

Patient Training and Guidance on How to Use the Technology

A total of 18% (3/17) of the studies reported that they conducted training for the pregnant women who participated in the intervention group. In the WeChat intervention, the research team taught the women how to use the app [33]. Homko et al [37] acknowledged the existing digital challenges and educated the women on how to use the technology. In a previous feasibility test, 15 women (47%) received computers, internet access, and a training session. Three of these women needed additional training. However, 22% of the women in the intervention group did not access the telemedicine system or use it. In the study limitations, the authors reported that, out of an average of 8 weeks of follow-up, the women used and transmitted their measurements on an average of only 3 weeks [37,44]. In the Pregnant+ app, researchers assumed that the women were capable of downloading and using the app, and no training was offered [25]. In the rest of the studies (11/17, 65%), the information was not reported (Table 3).

Plain Language

The Pregnant+ app recognized the different literacy levels of the women. Thus, they amended the information that the women received in line with their needs. Here, too, contents were checked using the Kreuter health message checklist to assess developers' use of plain language [43,45]. Regarding Habits-GDM, interviews were conducted after the study trial. A total of 6 out of 9 women (67%) said that the educational lessons were easy to understand [35]. In all other interventions, information regarding the level of language and medical jargon used was not available (Table 3).

Display and Organization—Simple Navigation

Navigation and screen display were described in 25% (3/12) of the interventions. In the Pregnant+ app, a graphic designer and language editor were involved in the design and development phases. The app content was designed and organized hierarchically and included only 4 icons to ease use and avoid overburdening the pregnant women [43]. Homko et al [44], in a preceding feasibility study, described the system's web screens for measurements and information, the data entry section, the sent questions, and the data appearance. Regarding the SineDie application, a previous study was conducted to evaluate the system. In the manuscript, the authors described the system design and architecture and included photos to demonstrate the view and drop-down lists that the women used when entering data as well as the summary of the electronic logbook [31,46]. In the Habits-GDM enrollment phase, those who were included in the study were required to be able to use a mobile phone and navigate through the application. The intervention interviews conducted showed that only 3 out of 9 users (33%) reported that all the information was in one place, and 3 out of 12 users (25%) said that they had difficulties in searching for features in the diet-tracking function while using the application [35,36]. A satisfaction survey conducted on the DiabeTIC web-based telemedicine platform found that the women's mean score on the item "understanding how to navigate through the platform" was 6.7 (SD 3.0, range 1-10), and the mean score on the item "adequate visualization of all information" was 8.2 (SD 1.8, range 4-10) [30,48]. Mackillop et al [49] tested the usability and reliability of the app and, following the results, app displays and colors were changed. In addition, to ease app navigation, they added illustrations on the screen icons [40,49] (Table 3).

Testing the Intervention

We found that 50% (6/12) of the interventions conducted preimplementation studies to examine user experience, ease of use, understanding of content, and app navigation following the prototype's design and development. Homko et al [44] tested the intervention's feasibility focusing on how well the pregnant women communicated with their health care provider and used the telemedicine system for better maternal and infant outcomes. The Pregnant+ app involved 21 pregnant women in its design and development phase, and 2 user-involvement studies were conducted afterward [43]. In addition, 2 qualitative studies were carried out to examine women's and providers' experiences and attitudes toward the Pregnant+ app [27,28]. Regarding the SineDie web-based clinical decision support system, a feasibility study was conducted. A total of 25 women participated in a validity study, and 90 women were randomized and participated in a clinical trial study testing effectiveness [46]. Mackillop et al [49], before the implementation of their intervention, conducted beta testing for system use (n=7) and the service

development phase (n=48). Pérez-Ferre et al [47] conducted a pilot study to test the telemedicine system's feasibility in clinical practice, reduction of face-to-face visits, and participants' satisfaction. The DiabeTIC pilot study examined 32 participants' satisfaction levels with the use of the telemedicine platform. A total of 7 of the 32 patients were women diagnosed with GDM [48]. In total, 25% (3/12) of the interventions evaluated the mHealth apps and telemedicine systems during implementation or afterward. In the WeChat intervention, short interviews were conducted, but no details on the questions and answers were described in the manuscript [32]. Regarding Habits-GDM, a qualitative study examined patients' experiences using the app after the study trial was delivered [36]. The TeleGDM web-based telemedicine study used mixed methods to examine patient and provider usability, acceptance, and satisfaction with using the technology [51] (Table 3).

Effectiveness

Overview

A summary of the outcome measurements and significant results of the 82% (14/17) of quantitative studies included in this review is provided in Table 4.

Table 4. Maternal and neonatal outcome measurements and significant results (N=14).

Study	Maternal clinical outcome measurements	Pregnant women's lifestyle outcome measurements	Neonatal outcome measurements	Significant outcomes

Tian et al [32], 2021	• Glycemic control:
	  • Number of BG[a] levels within the control range
	  • FBG[b] and 2hBG[c]
	• Maternal outcomes:
	  • PROM[d]
	  • Postpartum hemorrhage
	  • Delivery mode—CS[e], vaginal, or vacuum
		• N/A[f]	• Preterm birth
			• Birth weight
				• Differences in glycemic qualification, but clinical maternal and neonatal outcomes were not significantly different between the IG[g] and CG[h]

Yang et al [33], 2018	• Glycemic control:
	  • FBG
	  • 1hBG[i]
	  • 2hBG
	• Maternal outcomes:
	  • Delivery mode—CS, vaginal, or vacuum
	  • PIH[j]
	  • PROM
		• N/A	• Birth weight or macrosomia
			• Admission to the NICU[k]
			• Neonatal jaundice or hyperbilirubinemia
			• Neonatal hypoglycemia
			• Preterm birth
				• Glycemic control:
				  • FBG: P<.001 (IG vs CG)
				  • 2hBG: P<.001 (IG vs CG)
				• Maternal outcomes:
				  • Premature delivery (IG vs CG): P=.03
				  • CS was more likely in the IG: P=.03

Homko et al [37], 2012	• Glycemic control:
	  • FBG
	• Maternal outcomes:
	  • Delivery mode—CS, vaginal, or vacuum
	  • PIH or pre-eclampsia
	  • PROM
	  • Gestational week of delivery
		• Use of the system	• Admission to the NICU
			• Birth weight or macrosomia
			• Apgar score
			• Neonatal hypoglycemia
			• Respiratory morbidities
				• Women who used the internet sent more transmissions than women who used the phone or IVR[l] system: P=.007
				• Women with higher incomes transmitted more frequently: P<.01

Garnweidner-Holme et al [26], 2020	• Maternal outcomes: • Dietary changes	• N/A	• N/A	• No significant differences

Borgen et al [25], 2019	• Glycemic control:
	  • 2-hour glucose level postpartum OGTT[m]
	• Maternal outcomes:
	  • Delivery mode—CS, vaginal, or vacuum
		• Engagement with health	• Birth weight or macrosomia
			• Admission to the NICU
			• Apgar score
				• Women in the IG were less likely to have an emergency CS compared with the CG—overall mode of delivery: P=.03. However, when the women were stratified by parity, this difference was no longer statistically significant
				• Higher number of women reported that apps made them more engaged with their health: P<.01
				• However, a single self-constructed, nonvalidated question was used to measure this, and it was not specific to the intervention app

Study	Maternal clinical outcome measurements	Pregnant women's lifestyle outcome measurements	Neonatal outcome measurements	Significant outcomes
Pérez-Ferre et al [29], 2010	• Glycemic control: • Change in HbA1c • Maternal outcomes: • PIH or pre-eclampsia • Delivery mode—CS, vaginal, or vacuum • Weight gain • Gestational week of delivery • BP[n]	• Number of outpatient visits	• Birth weight or macrosomia • Admission to the NICU • Neonatal hypoglycemia • Preterm birth • Shoulder dystocia	• Reduction in outpatient clinic visits in women from the telemedicine group (P<.03) • The women in the IG had more contacts with health personnel and took up less time (P<.001) than those in the CG
Surendran et al [35], 2021	• N/A	• Frequency of: • Application use • Access to educational lessons • Coaching massages received	• N/A	• Only means, SDs, and percentage results
Carral, et al [30], 2015	• Glycemic control: • Change in HbA1c • Need for insulin • Maternal outcomes: • Delivery mode—CS, vaginal, or vacuum • PIH or pre-eclampsia • Gestational week of delivery • Weight gain • BP	• Number of patient visits: • GDU[o] • Obstetrics service • Emergency • General practitioner	• Birth weight or macrosomia • Neonatal hypoglycemia • Preterm birth	• Women in the IG required insulin therapy less frequently than women in the CG (P=.02) • Women in the IG had a lower number of visits to the GDU (P<.001), nurse educator (P<.001), and general practitioner (P<.001) than patients in the CG
Guo et al [34], 2019	• Glycemic control: • Change in HbA1c • 2-hour glucose level postpartum OGTT • Maternal outcomes: • Delivery mode—CS, vaginal, or vacuum • Gestational week of delivery • Weight gain	• Compliance • Number of outpatient visits	• Birth weight or macrosomia • Neonatal hypoglycemia • Shoulder dystocia	• Patient compliance was higher in the IG than in the CG (P<.001) • Frequency of outpatient service visits was lower in the IG compared with the CG (P<.001) • Weight gain in the IG was lower than in the CG (P<.001) • FBG (P<.001) and 2-hour postprandial (P<.001) were lower in the IG than in the CG
Mackillop et al [40], 2018	• Glycemic control: • Change in HbA1c • Longitudinal glycemic control • Maternal outcomes: • Delivery mode—CS, vaginal, or vacuum • PIH or pre-eclampsia • Weight gain • Gestational week of delivery	• Treatment satisfaction • Compliance	• Birth weight or macrosomia • Admission to the NICU • Neonatal jaundice or hyperbilirubinemia • Neonatal hypoglycemia • Shoulder dystocia	• Cesarean delivery was lower in the IG compared with the CG (P=.005), with notably fewer emergency cesarean deliveries in the IG • Women in the IG had higher satisfaction with care (P=.049) • Compliance with BG readings was better in the IG (OR[p] 2.44, 95% CI 1.29-4.61)
Miremberg et al [41], 2018				

Study	Maternal clinical outcome measurements	Pregnant women's lifestyle outcome measurements	Neonatal outcome measurements	Significant outcomes

	• Glycemic control:
	  • Longitudinal glycemic control
	  • Need for insulin
	• Maternal outcomes:
	  • Delivery mode—CS, vaginal, or vacuum
	  • PIH or pre-eclampsia
	  • Gestational week of delivery
	  • Polyhydramnios
	  • Perinatal tears
		• Compliance (the actual BG measurements vs instructed measurements)	• Birth weight or macrosomia
			• Admission to the NICU
			• Neonatal jaundice or hyperbilirubinemia
			• Neonatal hypoglycemia
			• Shoulder dystocia
			• Neonatal respiratory morbidity
			• Neonatal death
				• Compliance was higher in the IG than in the CG (P<.001)
				• Mean BG was lower in the IG than in the CG (P<.001)
				• Overall rate of insulin treatment was lower in the IG than in the CG (P=.04)
				• FBG (P<.001) and 1-hour postprandial (P<.001) were lower in the IG than in the CG

Yew et al [36], 2021	• Glycemic control:
	  • Longitudinal glycemic control
	  • Need for insulin
	• Maternal outcomes:
	  • Delivery mode—CS, vaginal, or vacuum
	  • PIH or pre-eclampsia
	  • Weight gain
	  • Gestational week of delivery
	  • Need for insulin
		• Mental and emotional health outcomes
		• Depression
		• Anxiety
		• Compliance
			• Birth weight or macrosomia
			• Admission to the NICU
			• Neonatal jaundice or hyperbilirubinemia
			• Neonatal hypoglycemia
			• Apgar score
			• Shoulder dystocia
			• Neonatal respiratory morbidity
			• Neonatal death
				• Glucose above the targets was significantly lower in the IG than in the CG (before meal: P=.003; 2 hours after meal: P=.001)
				• Overall, neonatal complications were lower in the IG (38.1%) than in the CG (53.7%; P=.006)

Rasekaba et al [38], 2018	• Glycemic control: • Longitudinal glycemic control • Insulin dose • Maternal outcomes: • Delivery mode—CS, vaginal, or vacuum	• Health service use	• Admission to the NICU • Macrosomia or infant or birth weight	• Women in the CG reached optimal glycemic control (maximum insulin dose) quicker than women in the IG (mean 4.3, SD 4.2 weeks vs mean 7.6, SD 4.5 weeks; P<.001) and had fewer insulin titrations (P=.04)

[a]BG: blood glucose.

[b]FBG: fasting BG.

[c]2hBG: 2-hour BG.

[d]PROM: premature rupture of membranes.

[e]CS: cesarean section.

[f]N/A: not applicable.

[g]IG: intervention group.

[h]CG: control group.

[i]1hBG: 1-hour BG.

[j]PIH: pregnancy-induced hypertension.

[k]NICU: neonatal intensive care unit.

[l]IVR: interactive voice response.

[m]OGTT: oral glucose tolerance test.

[n]BP: blood pressure.

[o]GDU: gestational diabetes unit.

[p]OR: odds ratio.

Glycemic Control

Interventions targeting pregnant women with GDM primarily focused on glycemic control. The apps and telemedicine interventions enabled the women to transmit their blood glucose measurements and receive feedback or alerts on their glucose values as well as treatment recommendations. A total of 79% (11/14) of the studies examined glycemic control following the intervention. In total, 14% (2/14) of the studies (WeChat and Guo et al [34]—Dnurse app) reported significant differences in fasting plasma glucose, 2-hour fasting blood glucose, 1-hour postprandial [33,34], and HbA1c before delivery [34]. In 14%

(2/14) of the studies (Miremberg et al [41] and Yew et al [36]—Habits-GDM), pregnant women's longitudinal mean blood glucose values were lower [36,41], and 21% (3/14) of the studies (Carral et al [30]—DiabeTIC, Rasekaba et al [38]—TeleGDM, and Miremberg et al [41]) reported a decrease in the need for insulin therapy in the intervention group.

Maternal Outcomes

A total of 79% (11/14) of the studies compared maternal outcomes between women who participated in the intervention and those in the control group receiving usual standard care. The delivery mode included CS, vaginal, or vacuum. Only 9% (1/11) of the studies reported significantly lower rates of cesarean delivery in the intervention group than in the control group (Mackillop et al [40]). Guo et al [34] found significant differences in weight gain in favor of the intervention group in comparison with the control group (Dnurse app).

Maternal Lifestyle Measurements

A total of 71% (10/14) of the studies evaluated women's lifestyle outcomes. These outcomes included engagement with their health condition, depression and anxiety, satisfaction, compliance, use of the system, and changes in the number of clinical visits. In the Pregnant+ app intervention, 84% of the women reported that the app increased engagement with their health compared with 64% in the control group (P<.01) [25]. Another study showed that women in the intervention group transmitted more glucose measurements [37]. A lower number of patient visits was reported in 30% (3/10) of these studies (Pérez-Ferre et al [29], Carral et al [30]—DiabeTIC, and Guo et al [34]—Dnurse app), and higher patient compliance following the intervention was found in 30% (3/10) of the studies (Guo et al [34]—Dnurse app**,** Mackillop et al [40], and Miremberg et al [41]). Overall, satisfaction with care was found to be significantly higher in only 10% (1/10) of the studies (Mackillop et al [40]).

Neonatal Outcomes

Neonatal outcomes were examined in 79% (11/14) of the studies. Only 9% (1/11) of these studies found a significantly lower difference in the composite overall neonatal complications (birth trauma, hypoglycemia, hyperbilirubinemia, respiratory distress, neonatal intensive care unit admission, and perinatal death) in the intervention group compared with the control group. However, no differences were found in each outcome individually (Yew et al [36]—Habits-GDM).

Discussion

Principal Findings

The use of mHealth apps and digital platforms as a resource for information and pregnancy follow-up among pregnant women is rising. However, digital mHealth interventions require consideration of users' cognitive and technical skills, education level, and digital literacy level and of cultural appropriateness in addressing dietary preferences and religious customs. These factors are crucial not only to enhance the use of an mHealth app but also to promote better outcomes for pregnant women from diverse ethnic and socioeconomically disadvantaged populations.

This review highlighted that, although existing digital technologies may improve glycemic control and diabetes self-management, there is a consistent deficit in assessing for cultural and digital health literacy appropriateness for pregnant women diagnosed with GDM. Only 17% (2/12) of the interventions addressed language diversity, dietary habits, and culturally appropriate recipes for their patients [36,43]. Only 25% (3/12) of the interventions received a score of ≥3 for digital health literacy appropriateness in our assessment [25,26,30,37]. Owing to the limited evidence that exists, it is hard to understand how participants' cultural customs and preferences, as well as their level of digital health literacy, affected the interventions' effectiveness or were associated with better maternal and infant health outcomes.

Assessment of digital health literacy and cultural needs is essential to identify obstacles and barriers to the adoption and use of mHealth and telemedicine systems and enhance the usability of any technology for health care. We found, for example, that only 17% (2/12) of the interventions trained patients on how to use the applications [33,37], and 6% (1/17) of the studies identified technological literacy as a barrier to using the service [50]. In 2021, the World Health Organization released its Global Strategy on Digital Health 2020-2025. The report emphasizes that, as digital health systems and interventions become more common, literacy will become a crucial determinant in the adoption of these technologies [54].

Many different measurements for digital and health literacy exist in the literature [55], but guidelines for the design and development of appropriate digital health interventions, especially for an audience with low literacy levels, are still scarce. In 2001, the US Department of Health and Human Services, Office of Minority Health, published the National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care to assist health care providers in becoming culturally competent and sensitive [56]. These 14 standards are a call to action and include, for example, language assistance services both verbally and in writing as well as training of staff. Thus, although there is increasing awareness of providing suitable care for patients with diverse values, beliefs, and behaviors, it is not yet part of the strategic planning included in the design of digital apps and technologies for women diagnosed with GDM.

It is important to also address the major limitation in GDM studies because of the lack of international consensus between countries (eg, the United States, Europe, and Australia) or health organizations (eg, the American Diabetes Association, American College of Obstetricians and Gynecologists, World Health Organization, European Board, and College of Obstetrics and Gynecology) on GDM screening methods and diagnosis threshold criteria. Different approaches exist, and there is no one standard treatment protocol [57,58]. Therefore, it is difficult to compare the effectiveness of therapies and mHealth and telehealth interventions between different countries because of the variety in the standards of care and the influence of local practices. Tsakirdis et al [59], in their review on national and

international guidelines for diagnosis and management of GDM, presented similarities and differences between countries and organizations. In their conclusions, they emphasized the need for future research to resolve guideline conflicts and provide 1 international standard protocol for the screening and management of women with GDM.

An additional limitation we observed in studies reporting on GDM mHealth apps relates to the study sample size. The studies included in this review did not report in the methods section on the power analysis calculation that they conducted to determine their study sample size. Moreover, many studies (5/13, 38%) reported on small intervention groups. A small sample size can result in an underpowered study, which can lead to biased conclusions regarding the effectiveness of the interventions.

Strengths and Limitations

This review has several strengths and limitations that need to be acknowledged. The framework we used provided us with the foundation for a rigorous and transparent method to conduct this scoping review. A comprehensive literature search was conducted, and the papers retrieved were screened and selected by 2 independent reviewers (YS and SS). Both reviewers met on a regular basis for discussions and to resolve disagreements. However, our search was restricted to the English language. Manuscripts that were published in other languages were excluded, and it could be that their interventions were relevant. An additional limitation may be a result of our search focusing on studies published between 2010 and 2021. Technologies evolve and change day by day. Computers and mobile phone generations continue to improve over time, and there may be a difference in our evaluation because of improved technological abilities over the years.

Conclusions and Future Directions

This review explored the published literature on digital interventions for GDM. Although studies on digital technologies for health self-management exist, this review found only 12 published interventions and fewer studies that evaluated and designed the technology for pregnant women diagnosed with GDM in accordance with the patients' cultural needs and digital health literacy levels. Thus, there is insufficient evidence regarding the effectiveness and benefits of mHealth and telemedicine systems for women from diverse backgrounds. Future research is needed to better understand how best to adapt and implement cultural and literacy factors in the design of digital technology for GDM management.

Acknowledgments

YB was supported by a postdoctoral scholarship from The Russell Berrie Galilee Diabetes SPHERE (Social Precision-medicine Health Equity Research Endeavour), Bar-Ilan University.

Authors' Contributions

YB and SS led the scoping review, contributed to all phases of the study design and analysis, and wrote the main body of the manuscript. All coauthors (YB, EY, YP, NS, and SS) contributed to the draft writing, editing, and discussion of the results and reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for Scoping Reviews (PRISMA-ScR): checklist and explanation.
[DOCX File , 29 KB - jmir_v24i10e37844_app1.docx]

Multimedia Appendix 2
Online search strategy conducted in October 2021.
[DOCX File , 30 KB - jmir_v24i10e37844_app2.docx]

References

1. Gestational diabetes. International Diabetes Federation. 2020 Jan 7. URL: https://www.idf.org/our-activities/care-prevention/gdm [accessed 2022-06-30]
2. Kc K, Shakya S, Zhang H. Gestational diabetes mellitus and macrosomia: a literature review. Ann Nutr Metab 2015;66 Suppl 2:14-20 [FREE Full text] [doi: 10.1159/000371628] [Medline: 26045324]
3. McGuane JT, Grlj L, Peek MJ. Obesity, gestational diabetes and macrosomia are associated with increasing rates of early-term induction of labour at The Canberra Hospital. Aust N Z J Obstet Gynaecol 2019 Apr 02;59(2):215-220. [doi: 10.1111/ajo.12820] [Medline: 29718540]
4. Saravanan P, Magee LA, Banerjee A, Coleman MA, Von Dadelszen P, Denison F, et al. Gestational diabetes: opportunities for improving maternal and child health. lancet Diabetes Endocrinol 2020 Sep;8(9):793-800. [doi: 10.1016/s2213-8587(20)30161-3]

5. Buhary B, Almohareb O, Aljohani N, Alzahrani S, Elkaissi S, Sherbeeni S, et al. Glycemic control and pregnancy outcomes in patients with diabetes in pregnancy: a retrospective study. Indian J Endocrinol Metab 2016;20(4):481-490 [FREE Full text] [doi: 10.4103/2230-8210.183478] [Medline: 27366714]
6. Parsons J, Sparrow K, Ismail K, Hunt K, Rogers H, Forbes A. Experiences of gestational diabetes and gestational diabetes care: a focus group and interview study. BMC Pregnancy Childbirth 2018 Jan 11;18(1):25 [FREE Full text] [doi: 10.1186/s12884-018-1657-9] [Medline: 29325518]
7. Berry D, Johnson Q, Stuebe A. Monitoring and managing mothers with gestational diabetes mellitus: a nursing perspective. Nurs Res Rev 2015 Oct:91. [doi: 10.2147/nrr.s64534]
8. Chen P, Wang S, Ji J, Ge A, Chen C, Zhu Y, et al. Risk factors and management of gestational diabetes. Cell Biochem Biophys 2015 Mar 01;71(2):689-694. [doi: 10.1007/s12013-014-0248-2] [Medline: 25269773]
9. Denney JM, Quinn KH. Gestational diabetes: underpinning principles, surveillance, and management. Obstet Gynecol Clin North Am 2018 Jun;45(2):299-314. [doi: 10.1016/j.ogc.2018.01.003] [Medline: 29747732]
10. Teitelman AM. Using mobile technologies to improve women's health. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):830-832. [doi: 10.1016/j.jogn.2018.09.004] [Medline: 30292776]
11. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
12. Kreps G. The relevance of health literacy to mHealth. ISU 2017 Jun 26;37(2):123-130 [FREE Full text] [doi: 10.3233/isu-170828]
13. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
14. Marciano L, Camerini A, Schulz PJ. The role of health literacy in diabetes knowledge, self-care, and glycemic control: a meta-analysis. J Gen Intern Med 2019 Jun 15;34(6):1007-1017 [FREE Full text] [doi: 10.1007/s11606-019-04832-y] [Medline: 30877457]
15. Guo SH, Hsing H, Lin J, Lee C. Relationships between mobile eHealth literacy, diabetes self-care, and glycemic outcomes in Taiwanese patients with type 2 diabetes: cross-sectional study. JMIR Mhealth Uhealth 2021 Feb 05;9(2):e18404 [FREE Full text] [doi: 10.2196/18404] [Medline: 33544088]
16. Craig L, Sims R, Glasziou P, Thomas R. Women's experiences of a diagnosis of gestational diabetes mellitus: a systematic review. BMC Pregnancy Childbirth 2020 Feb 07;20(1):76 [FREE Full text] [doi: 10.1186/s12884-020-2745-1] [Medline: 32028931]
17. Carolan M. Diabetes nurse educators' experiences of providing care for women, with gestational diabetes mellitus, from disadvantaged backgrounds. J Clin Nurs 2014 May;23(9-10):1374-1384. [doi: 10.1111/jocn.12421] [Medline: 24313229]
18. Broderick J, Devine T, Langhans E, Lemerise A, Lier S, Harris L. Designing health literate mobile apps. Institute of Medicine. 2014 Jan 28. URL: http://nam.edu/wp-content/uploads/2015/06/HealthLiterateApps.pdf [accessed 2022-01-30]
19. Chen Q, Carbone ET. Functionality, implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]
20. Duarte-Gardea MO, Gonzales-Pacheco DM, Reader DM, Thomas AM, Wang SR, Gregory RP, et al. Academy of nutrition and dietetics gestational diabetes evidence-based nutrition practice guideline. J Acad Nutr Diet 2018 Sep;118(9):1719-1742. [doi: 10.1016/j.jand.2018.03.014] [Medline: 29859757]
21. 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]
22. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
23. The Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2015 edition / Supplement. The Joanna Briggs Institute. 2015. URL: https://nursing.lsuhsc.edu/JBI/docs/ReviewersManuals/Scoping-.pdf [accessed 2021-09-21]
24. 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]
25. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
26. Garnweidner-Holme L, Henriksen L, Torheim LE, Lukasse M. Effect of the pregnant+ smartphone app on the dietary behavior of women with gestational diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 04;8(11):e18614 [FREE Full text] [doi: 10.2196/18614] [Medline: 33146620]
27. Garnweidner-Holme L, Hoel Andersen T, Sando MW, Noll J, Lukasse M. Health care professionals' attitudes toward, and experiences of using, a culture-sensitive smartphone app for women with gestational diabetes mellitus: qualitative study. JMIR Mhealth Uhealth 2018 May 14;6(5):e123 [FREE Full text] [doi: 10.2196/mhealth.9686] [Medline: 29759959]

28. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
29. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
30. Carral F, Ayala MD, Fernández J, González C, Piñero A, García G, et al. Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. Diabetes Technol Ther 2015 May;17(5):349-354. [doi: 10.1089/dia.2014.0223] [Medline: 25629547]
31. Albert L, Capel I, García-Sáez G, Martín-Redondo P, Hernando ME, Rigla M. Managing gestational diabetes mellitus using a smartphone application with artificial intelligence (SineDie) during the COVID-19 pandemic: much more than just telemedicine. Diabetes Res Clin Pract 2020 Nov;169:108396 [FREE Full text] [doi: 10.1016/j.diabres.2020.108396] [Medline: 32890548]
32. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881 [FREE Full text] [doi: 10.2196/22881] [Medline: 33783365]
33. Yang P, Lo W, He Z, Xiao X. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J Obstet Gynaecol Res 2018 Jul 23;44(7):1228-1234. [doi: 10.1111/jog.13669] [Medline: 29797375]
34. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
35. Surendran S, Lim CS, Koh GC, Yew TW, Tai ES, Foong PS. Women's usage behavior and perceived usefulness with using a mobile health application for gestational diabetes mellitus: mixed-methods study. Int J Environ Res Public Health 2021 Jun 21;18(12):6670 [FREE Full text] [doi: 10.3390/ijerph18126670] [Medline: 34205744]
36. Yew TW, Chi C, Chan S, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
37. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
38. Rasekaba TM, Furler J, Young D, Liew D, Gray K, Blackberry I, et al. Using technology to support care in gestational diabetes mellitus: quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
39. Khalil C. Understanding the adoption and diffusion of a telemonitoring solution in gestational diabetes mellitus: qualitative study. JMIR Diabetes 2019 Nov 28;4(4):e13661 [FREE Full text] [doi: 10.2196/13661] [Medline: 31778118]
40. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
41. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
42. Porta M. A Dictionary of Public Health. Oxford, United Kingdom: Oxford University Press; 2018.
43. Garnweidner-Holme L, Borgen I, Garitano I, Noll J, Lukasse M. Designing and developing a mobile smartphone application for women with gestational diabetes mellitus followed-up at diabetes outpatient clinics in Norway. Healthcare (Basel) 2015 May 21;3(2):310-323 [FREE Full text] [doi: 10.3390/healthcare3020310] [Medline: 27417764]
44. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
45. Kreuter M, Farrell DW, Olevitch LR, Brennan LK. Tailoring Health Messages Customizing Communication With Computer Technology. Mahwah, NJ: Lawrence Erlbaum Associates; 2013.
46. Caballero-Ruiz E, García-Sáez G, Rigla M, Villaplana M, Pons B, Hernando ME. A web-based clinical decision support system for gestational diabetes: automatic diet prescription and detection of insulin needs. Int J Med Inform 2017 Jun;102:35-49. [doi: 10.1016/j.ijmedinf.2017.02.014] [Medline: 28495347]

47. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, de la Cruz MJ, Martín P, et al. A telemedicine system based on internet and short message service as a new approach in the follow-up of patients with gestational diabetes. Diabetes Res Clin Pract 2010 Feb;87(2):e15-e17. [doi: 10.1016/j.diabres.2009.12.002] [Medline: 20044162]
48. Carral San Laureano F, Ayala Ortega MD, Jiménez Millán AI, Piñero Zaldivar A, García Calzado C, Prieto Ferrón M, et al. DiabeTIC website: a pilot study of satisfaction and impact on metabolic control. Endocrinología y Nutrición (English Edition) 2013 Oct;60(8):441-446. [doi: 10.1016/j.endoen.2013.10.016]
49. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: 10.1177/1932296814542271] [Medline: 25004915]
50. Rasekaba T, Nightingale H, Furler J, Lim WK, Triay J, Blackberry I. Women, clinician and IT staff perspectives on telehealth for enhanced gestational diabetes mellitus management in an Australian rural/regional setting. Rural Remote Health 2021 Jan 22:5983. [doi: 10.22605/rrh5983]
51. Rasekaba TM, Lim K, Blackberry I, Gray K, Furler J. Telemedicine for gestational diabetes mellitus (TeleGDM): a mixed-method study protocol of effects of a web-based gdm support system on health service utilization, maternal and fetal outcomes, costs, and user experience. JMIR Res Protoc 2016 Aug 09;5(3):e163 [FREE Full text] [doi: 10.2196/resprot.6044] [Medline: 27507708]
52. Doak CC, Doak LG, Root JH. Teaching patients with low literacy skills. Am J Nurs 1996;96(12):16M. [doi: 10.1097/00000446-199612000-00022]
53. Mackillop LH, Bartlett K, Birks J, Farmer AJ, Gibson OJ, Kevat DA, et al. Trial protocol to compare the efficacy of a smartphone-based blood glucose management system with standard clinic care in the gestational diabetic population. BMJ Open 2016 Mar 17;6(3):e009702 [FREE Full text] [doi: 10.1136/bmjopen-2015-009702] [Medline: 26988348]
54. Global Strategy on Digital Health 2020-2025. Geneva: World Health Organization; 2021.
55. 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]
56. National standards for culturally and linguistically appropriate services in health care. U.S. Department of Health and Human Services Office of Minority Health. 2001 Mar. URL: https://minorityhealth.hhs.gov/assets/pdf/checked/finalreport.pdf [accessed 2022-01-30]
57. Petrović O, Belci D. A critical appraisal and potentially new conceptual approach to screening and diagnosis of gestational diabetes. J Obstet Gynaecol 2017 Aug 03;37(6):691-699. [doi: 10.1080/01443615.2017.1306692] [Medline: 28467229]
58. Agarwal M. Consensus in gestational diabetes MELLITUS: looking for the holy grail. J Clin Med 2018 May 28;7(6):123 [FREE Full text] [doi: 10.3390/jcm7060123] [Medline: 29843401]
59. Tsakiridis I, Giouleka S, Mamopoulos A, Kourtis A, Athanasiadis A, Filopoulou D, et al. Diagnosis and management of gestational diabetes mellitus: an overview of national and international guidelines. Obstet Gynecol Surv 2021 Jun;76(6):367-381. [doi: 10.1097/OGX.0000000000000899] [Medline: 34192341]

Abbreviations

GDM: gestational diabetes mellitus
mHealth: mobile health
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
RCT: randomized controlled trial

Edited by T Leung; submitted 13.03.22; peer-reviewed by T Rasekaba, R Halkes; comments to author 02.06.22; revised version received 03.07.22; accepted 31.07.22; published 14.10.22.

Please cite as:
Birati Y, Yefet E, Perlitz Y, Shehadeh N, Spitzer S
Cultural and Digital Health Literacy Appropriateness of App- and Web-Based Systems Designed for Pregnant Women With Gestational Diabetes Mellitus: Scoping Review
J Med Internet Res 2022;24(10):e37844
URL: https://www.jmir.org/2022/10/e37844
doi:10.2196/37844
PMID:36240008

Review

Digital Mental Health Interventions for Depression: Scoping Review of User Engagement

Jessica M Lipschitz[1,2], PhD; Rachel Van Boxtel[1], BA; John Torous[2,3], MD; Joseph Firth[4,5], PhD; Julia G Lebovitz[1], BA; Katherine E Burdick[1,2], PhD; Timothy P Hogan[6,7], PhD

[1]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, United States

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[3]Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, United States

[4]Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

[5]Greater Manchester Mental Health NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom

[6]Center for Healthcare Organization and Implementation Research, Veterans Affairs Bedford Healthcare System, Bedford, MA, United States

[7]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Jessica M Lipschitz, PhD
Department of Psychiatry
Brigham and Women's Hospital
221 Longwood Ave
Boston, MA, 02115
United States
Phone: 1 617 732 6548
Email: jlipschitz@bwh.harvard.edu

Abstract

Background: While many digital mental health interventions (DMHIs) have been found to be efficacious, patient engagement with DMHIs has increasingly emerged as a concern for implementation in real-world clinical settings. To address engagement, we must first understand what standard engagement levels are in the context of randomized controlled trials (RCTs) and how these compare with other treatments.

Objective: This scoping review aims to examine the state of reporting on intervention engagement in RCTs of mobile app–based interventions intended to treat symptoms of depression. We sought to identify what engagement metrics are and are not routinely reported as well as what the metrics that are reported reflect about standard engagement levels.

Methods: We conducted a systematic search of 7 databases to identify studies meeting our eligibility criteria, namely, RCTs that evaluated use of a mobile app–based intervention in adults, for which depressive symptoms were a primary outcome of interest. We then extracted 2 kinds of information from each article: intervention details and indices of DMHI engagement. A 5-element framework of minimum necessary DMHI engagement reporting was derived by our team and guided our data extraction. This framework included (1) recommended app use as communicated to participants at enrollment and, when reported, app adherence criteria; (2) rate of intervention uptake among those assigned to the intervention; (3) level of app use metrics reported, specifically number of uses and time spent using the app; (4) duration of app use metrics (ie, weekly use patterns); and (5) number of intervention completers.

Results: Database searching yielded 2083 unique records. Of these, 22 studies were eligible for inclusion. Only 64% (14/22) of studies included in this review specified rate of intervention uptake. Level of use metrics was only reported in 59% (13/22) of the studies reviewed. Approximately one-quarter of the studies (5/22, 23%) reported duration of use metrics. Only half (11/22, 50%) of the studies reported the number of participants who completed the app-based components of the intervention as intended or other metrics related to completion. Findings in those studies reporting metrics related to intervention completion indicated that between 14.4% and 93.0% of participants randomized to a DMHI condition completed the intervention as intended or according to a specified adherence criteria.

Conclusions: Findings suggest that engagement was underreported and widely varied. It was not uncommon to see completion rates at or below 50% (11/22) of those participants randomized to a treatment condition or to simply see completion rates not reported at all. This variability in reporting suggests a failure to establish sufficient reporting standards and limits the conclusions

that can be drawn about level of engagement with DMHIs. Based on these findings, the 5-element framework applied in this review may be useful as a minimum necessary standard for DMHI engagement reporting.

(J Med Internet Res 2022;24(10):e39204) doi:10.2196/39204

KEYWORDS

mHealth; mobile apps; engagement; adherence; randomized controlled trials; depression

Introduction

Digital mental health interventions (DMHIs) are a promising avenue for accessible treatment for people with widespread and debilitating mental health issues such as depression. The field of psychiatry continues to struggle with an insufficient supply of highly trained providers able to offer evidence-based services who are accessible in terms of location and cost. While face-to-face, evidence-based psychotherapy remains the first-line treatment option for mild to moderate depression [1], emerging literature on DMHIs suggests that these too could be an effective stand-alone or supplemental treatment option [2,3]. These interventions have, therefore, generated significant public interest as they are more accessible and lower cost than face-to-face psychotherapy.

As interest has mounted, however, so too have concerns about low patient engagement with these interventions. In the last 10 years, several large implementation studies of DMHIs have shown that the majority of patients offered these interventions do not engage at the recommended frequency or complete the full course of treatment [4-6]. In a large implementation study, Gilbody et al [7] concluded that "while [DMHIs] have been shown to be efficacious in developer led trials, [they were] not effective in usual NHS care settings. The main reason for this was low adherence and engagement with treatment rather than lack of efficacy." Such low engagement rates threaten the clinical viability of these treatments.

DMHI engagement has been defined as a patient's initial adoption and sustained interactions with an intervention [8-10]. Within the broader construct of engagement, intervention adherence refers to the extent to which participants engage in the content of the intervention as intended. In the context of randomized controlled trials (RCTs) intervention adherence can be reported as the number of intervention completers with the criteria for completion being clearly specified. However, within the broader construct of engagement, other metrics, such as the rate at which participants download and initiate intervention use (ie, uptake), degree or level of use of the intervention, and duration of use of the intervention are also relevant.

Engagement is particularly important to consider in RCTs because low intervention engagement poses a threat to the validity of conclusions drawn. It could lead to underestimating the intervention effect especially if a dose-response relationship exists [11]. Furthermore, as discussed by Eysenbach [12], if a participant did not significantly engage with an intervention, it is difficult to conclude that the intervention produced a positive outcome even if such outcomes were observed. In these cases, we are left with questions about the extent to which confounding variables, such as attention from study staff, could have produced any observed intervention effect. Finally, when degree of intervention engagement is not clearly described in manuscripts, we lose information on how an intervention must be used to achieve observed effects. For example, if an 8-week intervention period was studied and a positive intervention effect was observed, but 70% of participants only used the intervention for the first 2 weeks of the intervention period, we may conclude that just 2 weeks of use may be producing positive results. Alternatively, we may conclude that a certain level of effect could be expected after 2 weeks of use, whereas a different, perhaps more pronounced effect, could be expected after 8 weeks of use.

The concept of what constitutes sufficient engagement with DMHIs is inherently messier than for some other types of mental health interventions. For example, sufficient engagement with antidepressant medications typically means taking a daily pill. In psychotherapy, sufficient engagement is typically defined as attending all planned psychotherapy sessions. Use of medication and appointment attendance are clear quantitative adherence metrics. In the case of DMHIs, however, heterogeneity in intervention design leaves us with considerably less clarity on appropriate intervention adherence metrics. Some DMHIs, such as the Get Happy Program [13], consist of a series of lessons or modules that are designed to be completed in a sequential fashion over a specified number of weeks. These programs mirror face-to-face therapy programs where there is an assumption of some established weekly content review or dedicated time commitment. Other DMHIs are designed to be used more frequently for briefer periods. For example, IntelliCare [14] is designed to be used on a daily basis, but length of time in the app is not prescribed. Still, other interventions (eg, the MONARCA System [15]) consist primarily of symptom monitoring and are designed to be used frequently to inform and support clinician-based care.

This inherent heterogeneity of DMHIs makes engagement difficult to compare across studies. It also calls for consideration of what constitutes appropriate reporting related to both the larger construct of engagement and the narrower construct of adherence. To date, reviews and meta-analyses related to engagement with DMHIs have tended to focus on related, but distinct concepts. For example, study dropout or study attrition has been evaluated as a proxy for treatment dropout, with findings suggesting significant dropout [16-18]. Similarly, user-rated acceptability and feasibility have been evaluated as proxies for engagement [19]. Finally, several recent reviews have explored variables related to user engagement with DMHIs [8,9,18]. However, to date, no review to our knowledge has explored the actual level of user engagement in RCTs of DMHIs. Therefore, the objective of this scoping review was to examine reporting on user engagement in RCTs of mobile app–based

interventions for symptoms of depression. Specifically, we sought to identify (1) the extent to which key engagement metrics are routinely reported and (2) what the metrics that are reported reflect about standard levels of engagement.

Methods

The creation of this report was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Extension for Scoping Reviews (Multimedia Appendix 1) [20].

Information Sources and Search Strategy

A systematic search was conducted using OvidSP to search 7 electronic databases, MEDLINE, Embase, Cochrane Central Register of Controlled Trials, Allied and Complementary Medicine, Health Management Information Consortium, Health Technology Assessment, and PsycINFO, for articles published through May 1, 2020 (Table 1). The search was conducted on May 7, 2020. In brief, the search strategy combined synonyms for the population of interest (patients with mental illness), the intervention modality (mobile phone apps), and the type of study (RCT). Search results were limited to the English language and studies of humans.

Table 1. Search strategy as used in OvidSP on May 7, 2020.

Search category	Search terms
Population	"depression" OR "depressive" OR "mental illness" OR "mental health" OR "mood disorder" OR "affective disorder" OR "anxiety" OR "panic disorder" OR "phobia" OR "bipolar" OR "psychosis" OR "schizophr*" AND
Intervention	"smartphone*" OR "mobile phone*" OR "cell phone*" OR "iphone" OR "android" OR "mhealth" OR "mobile application" OR "phone application" AND
Type of study	"randomised" OR "randomized" OR "randomly" OR "random assignment" OR "controlled trial" OR "clinical trial" OR "control group" OR "intervention"
Platform used	OvidSP
Databases selected for search	Ovid MEDLINE, Embase, Cochrane Central Register of Controlled Trials, Allied and Complementary Medicine, Health Management Information Consortium, Health Technology Assessment, and PsycINFO

Inclusion and Exclusion Criteria

Only articles published in peer-reviewed journals were included. Articles were deemed eligible if they were RCTs of mobile app–based interventions targeting adults (aged >18) with clinical depression, in which depressive symptoms were a primary outcome of interest, and retention in posttreatment assessments was reported. We defined a mobile app–based intervention as one that required use of a mobile device app as part of the treatment.

We defined studying a "clinically depressed" sample as meeting at least one of the following criteria: (1) eligibility criteria requiring participants to have scores on a depression self-report measure over an established clinical cutoff; (2) eligibility criteria requiring participants to have a psychiatric diagnosis per their medical record or per a structured clinical interview; or (3) reported average baseline scores on a depression self-report measure above an established clinical cutoff in all groups. When there was ambiguity on the established clinical cutoff for a self-report measure, we used the lowest published cutoff score.

At least two independent reviewers judged article eligibility (JML, JGL, or RVB), with any disagreements resolved through mediation with a third reviewer (TPH). The screening process began with title and abstract review followed by a full-text review of any articles that appeared potentially relevant based on the abstract/title review or where there was insufficient information in the abstract to determine eligibility.

Data Extraction and Synthesis

Data extraction occurred in 3 parts. First, data were extracted by one author (JGL or RVB). Next, the rationale for each datapoint and where it came from in the original articles were reviewed with JML. Finally, all datapoints considered ambiguous or disagreements between the authors who completed the initial data extraction and JML were reviewed with one additional author (TPH).

Two kinds of information were extracted from each article. First, intervention details were extracted, including the (1) clinical population, (2) length of the treatment period, (3) a description of the study conditions, (4) total sample size in each condition, and (5) whether human support by a coach or licensed clinician was offered as part of the intervention.

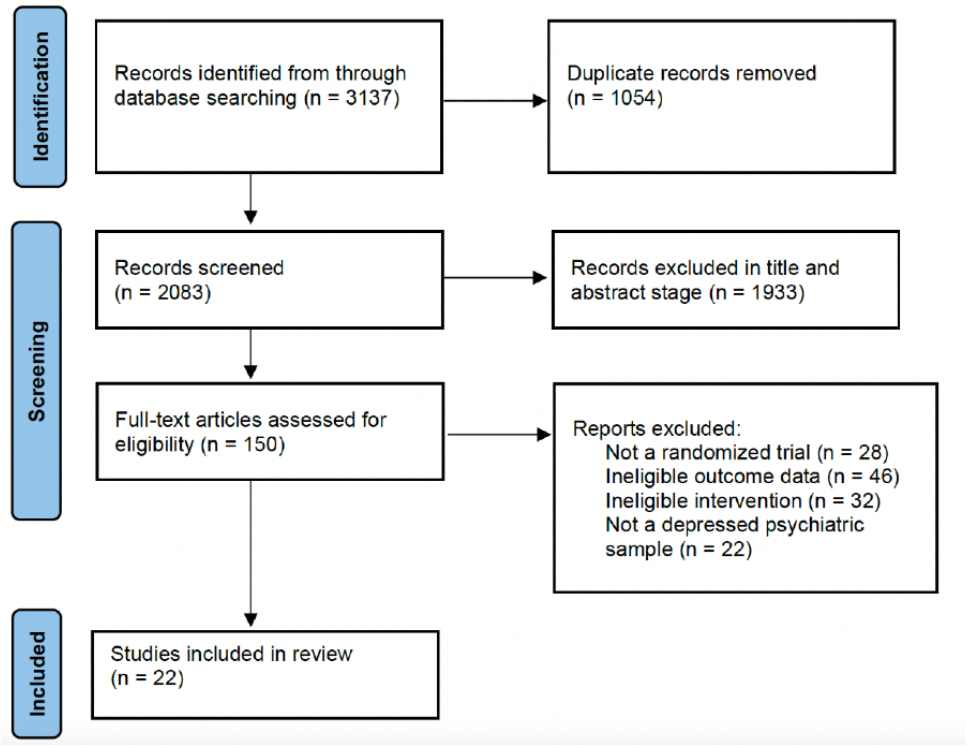
Second, a 5-element framework of minimum necessary DMHI engagement reporting, developed by our study team, was used to extract key descriptive and numeric indices of participant engagement with the intervention. Elements in this framework were as follows: (1) recommended intervention app use as communicated to participants at enrollment and, when reported, intervention app adherence criteria; (2) rate of uptake, defined as the number and percentage of participants randomized to the intervention who engaged with their assigned app at all; (3) level of intervention app use metrics, specifically number of times participants used the app and amount of time participants spent in the app; (4) duration of intervention app use metrics (ie, whether weekly use patterns were reported and the number and percentage of participants who used the app in the final week of the intervention period); and (5) number and percentage of participants randomized who could be considered intervention completers. Furthermore, for context, we identified whether studies used backend data or other methods (such as self-report) to quantify app usage and extracted any additional data presented on intervention engagement.

Results

Selection and Inclusion of Studies

The full systematic search retrieved a total of 3137 records (Figure 1). Following the removal of duplicate articles across electronic databases, 2083 articles were screened at the title-and-abstract phase. This identified 150 articles as potentially eligible, which were subsequently screened in full. Full-text screening resulted in the exclusion of 128 articles for reasons specified in Figure 1. A total of 22 independent studies [13,15,21-40] were ultimately eligible for inclusion.

Figure 1. PRISMA Search Diagram.



Characteristics of Included Studies

Detailed study characteristics are presented in Table 2. While all 22 studies included a clinically depressed sample and symptoms of depression as a primary outcome, the target populations differed. Of the 22 eligible studies, the following target populations were recruited: depression (n=13), suicidal ideation (n=1); depression or anxiety (n=3); bipolar disorder (n=1); medical population with clinically significant symptoms of depression (n=2); community sample (n=1); and college students (n=1). Intervention periods ranged from 2 weeks to 6 months and sample sizes ranged from 30 to 720. Interventions evaluated included a range of human support: 11 were entirely self-help interventions involving no human support, 9 involved a licensed clinician, 1 involved a clinical coach, and 1 included clinical support from research staff for whom licensure status was not specified. For descriptive purposes, apps studied were assigned to 1 of 3 categories: those intended to be used as daily self-management/skill-building tools (n=13); those intended to provide support in the context of clinician-administered care or to facilitate communication with clinicians (n=5); and treatments involving a discrete number of lessons/modules typically to be completed on a weekly basis (n=4).

Table 2. Study characteristics.

Study	Clinical population	Treatment period	Conditions	Sample size	Human contact	App category
Arean et al [21]	• Depression	• 4 weeks	• Project: EVO (gamified cognitive training app) • iPST (problem-solving therapy app) • Health Tips (app providing daily tips for improved health; control)	• 209 • 211 • 206	• None • None • None	Daily self-management/skill building
Bakker et al [22]	• Community sample	• 30 days	• MoodKit (CBT[a]-based app with a variety of tools) • MoodPrism (self-monitoring mood-tracking app) • MoodMission (CBT-based app that recommends CBT strategies in response to user-reported low moods and anxious feelings) • Waitlist (control)	• 78 • 78 • 78 • 78	• None • None • None • None	Daily self-management/skill building
Birney et al [23]	• Depression	• 6 weeks	• MoodHacker (CBT-based depression management app based on the "Coping with Depression" program) • Alternate care group (emailed links to 6 websites with depression information; control)	• 150 • 150	• None • None	Daily self-management/skill building
Borjalilu et al [24]	• College students	• 20 days • 6 weeks • 6 weeks	• Aramgar stress management app (mindfulness-based stress reduction) • Blended (Aramgar app for 20 days + 6 weeks of face-to-face therapy) • Face-to-face therapy only	• 20 • 28 • 20	• None • Clinician • Clinician	Daily self-management/skill building
Dahne et al [25]	• Depression	• 8 weeks	• ¡Apívate! (Spanish language brief behavioral activation mobile app) • iCouch CBT (Spanish language CBT mobile app; active control) • TAU[b] (control)	• 22 • 9 • 11	• None • None • N/A[c]	Daily self-management/skill building
Dahne et al [26]	• Depression	• 8 weeks	• Moodivate (brief behavioral activation mobile app) • MoodKit (CBT mobile app; active control) • TAU (control)	• 24 • 19 • 9	• None • None • N/A	Daily self-management/skill building
Faurholt-Jepsen et al [15]	• Bipolar disorder	• 6 months	• MONARCA system (daily self-monitoring app with feedback from clinician) • Placebo MONARCA (Android cellphone and TAU; control)	• 39 • 39	• Clinician • Clinician	Support for appointments/interaction with clinician
Fitzpatrick et al [27]	• Depression or anxiety	• 2 weeks	• Woebot (CBT-oriented conversational agent app) • "Depression in College Students" eBook created by the National Institute of Mental Health (informational booklet; control)	• 34 • 36	• None • None	Daily self-management/skill building
Guo et al [28]	• Patients with HIV with depression symptoms	• 3 months	• Run4Love (WeChat-based cognitive behavioral stress management course plus physical activity promotion) • Usual care (brochure on nutrition and usual care for HIV; control)	• 150 • 150	• Study staff (unclear if coach/clinician) • None	Discrete number of lessons/modules
Lüdtke et al [29]	• Depression	• 4 weeks	• Be Good to Yourself app (40 self-help strategies and exercises, based on CBT) • Waitlist (control)	• 44 • 44	• None • None	Daily self-management/skill building

Study	Clinical population	Treatment period	Conditions	Sample size	Human contact	App category
Ly et al [30]	• Depression	• 8 weeks	• Behavioral activation smartphone app • Mindfulness smartphone app	• 40 • 41	• Clinician • Clinician	Support for appointments/interaction with clinician[d]
Ly et al [31]	• Depression	• 9 weeks • 10 weeks	• Blended treatment (4 face-to-face behavioral activation sessions plus a smartphone app for support and suggestions between sessions) • Full behavioral activation (10 face-to-face behavioral activation sessions; control)	• 46 • 47	• Clinician • Clinician	Support for appointments/interaction with clinician
Mantani et al [32]	• Depression	• 9 weeks	• Kokoro app (8 sessions, CBT-based self-help app + antidepressant switch) • Antidepressant switch (control)	• 81 • 83	• Clinician • Clinician	Discrete number of lessons/modules
Moberg et al [33]	• Depression or anxiety	• 1 month	• Pacifica (guided CBT-based self-help app) • Waitlist (control)	• 253 • 247	• None • None	Daily self-management/skill building
Mohr et al [34]	• Depression or anxiety	• 8 weeks	• IntelliCare platform with coach (12 apps, each focusing on a single psychological or behavioral strategy) • IntelliCare platform self-guided • IntelliCare platform with recommendations • IntelliCare platform no recommendations	• 150[e] • 151 • 149 • 152	• Coached • None • Half coached • Half coached	Daily self-management/skill building
Motter et al [35]	• Depression	• 8 weeks	• Executive function/processing speed focused–computerized cognitive training • Verbal ability–focused computerized cognitive training	• 25 • 21	• None • None	Daily self-management/skill building
O'Toole et al [36]	• Patients referred for suicidal thoughts	• About 8 weeks (range 8-16 weeks, at clinician discretion)	• LifeApp'tite mobile app (suicide prevention app provided alongside suicide prevention psychotherapy protocol) • TAU (suicide prevention psychotherapy protocol; control)	• 60 • 69	• Clinician • Clinician	Support for appointments/interaction with clinician
Place et al [37]	• Primary care behavioral health patients	• 6 months	• Usual care (in behavioral health clinic) + Cogito's mobile sensing platform • Usual Care (control)	• 35 • 33	• Clinician • Clinician	Support for appointments/interaction with clinician
Proudfoot et al [40]	• Depression	• 7 weeks	• MyCompass Intervention (app with 12 skill-building modules derived from CBT, interpersonal psychotherapy, problem-solving therapy, and positive psychology) • Attention control (weekly mental health fact sheet delivered to email inbox; active control) • Waitlist (control)	• 242 • 248 • 230	• None • None • None	Discrete number of lessons/modules
Roepke et al [38][f]	• Depression	• 4 weeks	• CBT-PPT SB (SuperBetter game–like app with additional content from cognitive behavioral therapy and positive psychotherapy) • General SB (SuperBetter game–like app with additional content focused on self-esteem and acceptance) • Waitlist (control)	• 93 • 97 • 93	• None • None • N/A	Daily self-management/skill building

Study	Clinical population	Treatment period	Conditions	Sample size	Human contact	App category
Stiles-Shields et al [39]	• Depression	• 6 weeks	• Boost Me (an app intervention based on activity scheduling) • Thought Challenger (an app intervention based on thought restructuring) • Waitlist control (control)	• 10 • 10 • 10	• Cliniciang • Clinician • N/A	Daily self-management/skill building
Watts et al [13]	• Depression	• 8 weeks	• Get Happy Program mobile app (6 lessons on how to manage depression symptoms) • Get Happy Program computer delivered	• 22 • 30	• Clinician • Clinician	Discrete number of lessons/modules

[a]CBT: cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]N/A: no treatment administered.

[d]The intervention in Ly et al [30] contained elements of daily self-management/skill building, but completion was defined by interactions with a clinician so this was deemed primarily an intervention to support appointments/interaction with a clinician.

[e]Mohr et al [34] was a 2 × 2 factorial trial design. Group sample sizes specified here are not mutually exclusive.

[f]Roepke et al [38] reported that the SuperBetter intervention was targeted to occur on the iPhone, but could be used via a website on computers. This study was deemed eligible because the intention was for it to be smartphone based.

[g]Stiles-Shields et al [39] involved coaching, but is categorized as involving a clinician (not a coach) because the coach was a licensed clinician.

Reporting on Participant Engagement

Data extracted based on our 5-element framework are presented in Table 3 (with additional details presented in Multimedia Appendix 2). With the exception of Ludtke et al [29], all studies that reported on app usage indicated using backend data from the app to monitor app usage in the test condition(s). Ludtke et al [29] only offered self-reported app usage data; 14/22 papers (64%) reported the rate of app uptake defined as the number of participants randomized to the intervention condition(s) who engaged with the app at least once. Findings in those studies reporting the rate of app uptake indicated that between 42% and 100% of those participants randomized to an app-based DMHI condition engaged with the app at least once.

With regard to ongoing use, reports were varied. A total of 13 papers (59%) reported a level-of-use metric. The most common level-of-used metric was number of sessions/launches (n=12). Time spent in the app was a less popular level-of-use metric (n=4). Fewer papers reported metrics on duration of use. Only 5 studies (23%) reported weekly use patterns over the course of the intervention and the number of participants who were still using the intervention during the last week of the treatment period.

With regard to questions of whether participants completed the intervention as intended, reporting was also varied. Table 3 describes the app intervention instructions given to participants and app adherence criteria to the extent that these were specified in each article. Only 3 studies clearly reported the number of participants randomized to the DMHI who were considered to have completed the app-based components of the intervention as intended per specified intervention instructions. An additional 4 studies (footnote i in Table 3) reported the number of participants who met a specified adherence threshold such as using the intervention app once per week; 4 more studies reported metrics related to intervention completion, including percentage of patients who used the app on a daily basis (n=1; footnote m); percentage of patients completing the intervention based on a criterion defined by clinician contact rather than app use (n=2; marked by footnote o); and percentage of participants who downloaded all the intervention content (n=1; marked by footnote t). Findings in those studies reporting metrics related to intervention completion indicated that between 14.4% and 93.0% of participants randomized to a DMHI condition completed the intervention as intended or according to a specified app adherence criteria. Among the 11 studies reporting this metric, 6 reported that less than or equal to 50% of participants completed the intervention.

Table 3. Treatment engagement metrics for digital mental health interventions.[a]

Study and intervention name	App use instructions or adherence criteria	Rate of uptake[b], n (%)	Level of use		Duration of use		Completers[c], n (%) or %
			App uses	Minutes spent using the app, mean (SD)	Reported weekly use pattern	Used in the final week, n (%) or %	
Arean et al [21]							
Project: EVO	Use app 6 times/week for 30 minutes/day (3 or more times/week considered adherent)	177 (42.1)[d]	Mean 10.78 (SD 11.4)[e]	NR[f]	Yes	42 (20.1)[g]	30 (14.4)[g,h,i]
iPST	Use app as often as possible (1 or more times/week considered adherent)	—[j]	—[j]	NR	Yes	40 (19.0)[g]	40 (19.0)[g,h,i]
Health Tips App	No specific instructions, but daily advice was provided	NR	NR	NR	No	NR	NR
Bakker et al [22]							
MoodKit	No specific instructions reported	NR	NR	NR	No	NR	NR
MoodPrism	No specific instructions reported	NR	NR	NR	No	NR	NR
MoodMission	No specific instructions reported	NR	NR	NR	No	NR	NR
Birney et al [23]							
MoodHacker	Daily app use	NR	Mean 16.0 (SD 13.3)	78 (78)	No	NR	NR
Borjalilu et al [24]							
Aramgar app	Complete recommended app exercises daily	NR	NR	NR	No	NR	NR
Aramgar app with face-to-face therapy	Twice/week face-to-face workshops plus daily app exercises	NR	NR	NR	No	NR	NR
Dahne et al [25]							
¡Apívate!	Use app once/day (1 or more times/week considered adherent)	22 (100)	Mean 61.4 (SD 91.7)	65.8 (82.8)	Yes	11 (50)	11 (50)[h,i]
iCouch CBT	Use app once/day (1 or more times/week considered adherent)	NR	NR	NR	Yes	33[g]	33[g,h,i]
Dahne et al [26]							
Moodivate	Use the app once/day (1 or more times/week considered adherent)	21 (100)[k]	Mean 46.8 (SD 30.1)	120.8 (101.0)	Yes	9 (50)[k]	9 (50)[h,i,k]
MoodKit	Use app once/day	NR	NR	NR	No	NR	NR
Faurholt-Jepsen et al [15]							
MONARCA	Use app for self-monitoring daily	34 (87.2)[l]	NR	NR	No	NR	93.0[m]
Fitzpatrick et al [27]							
Woebot	Daily monitoring and “regular check-ins”	34 (100)	Mean 12.14 (SD 2.23)	NR	No	NR	NR
Guo et al [28]							
Run4Love	Complete 9 cognitive behavioral stress management sessions, 3 review sessions, and set weekly physical activity goal	NR	NR	NR	No	NR	NR
Lüdtke et al [29]							
Be Good to Yourself app	Use app “several times a week”	26 (59.1)[n]	NR	NR	No	NR	19 (43.2)[n]

Study and intervention name	App use instructions or adherence criteria	Rate of uptake[b], n (%)	Level of use		Duration of use		Completers[c], n (%) or %
			App uses	Minutes spent using the app, mean (SD)	Reported weekly use pattern	Used in the final week, n (%) or %	
Ly et al [30]							
Behavioral activation smartphone app	Add at least two behavioral goals to the app and register/write a reflection in the app when these goals were completed	81 (96.4)[d]	NR	NR	No	NR	25 (63.0)[e,o]
Mindfulness smartphone app	Use audio tracks with exercises to facilitate the practice of mindfulness	—[j]	—[j]	NR	No	NR	32 (78.0)[e,o]
Ly et al [31]							
Blended treatment	No specific instructions reported	NR	NR	NR	No	NR	42 (91.3)[o]
Mantani et al [32]							
CPT-Kokoro app	Complete 8 mobile app sessions, 1 per week	80 (98.76)	Mean 7.01 (SD 1.5)[p]	NR	Yes	43 (53.1)	43 (53.1)[p]
Moberg et al [33]							
Pacifica	No specific instructions reported	246 (97.2)	Median 19 (range 1-286)[e]	NR	No	NR	NR
Mohr et al [34]							
IntelliCare: coached	No specific instructions reported (last app use at or after 7 weeks considered adherent)	143 (95.3)[l]	Median 215 (IQR 141-330.8)	NR	Yes	136 (90.7)[q]	136 (90.7)[m,p]
IntelliCare: self-guided	No specific instructions reported (last app use at or after 7 weeks considered adherent)	151 (100)[l]	Median 218 (IQR 113-310)	NR	Yes	126 (83.4)[q]	126 (83.4)[m,p]
IntelliCare: recommendations	No specific instructions reported (last app use at or after 7 weeks considered adherent)	146 (98.0)[l]	Median 232 (IQR 126-356)	NR	Yes	132 (88.6)[q]	132 (88.6)[m,p]
IntelliCare: no recommendations	No specific instructions reported (last app use at or after 7 weeks considered adherent)	148 (97.4)[l]	Median 201.5 (IQR 125.8-285.5)	NR	Yes	130 (85.5)[q]	130 (85.5)[m,p]
Motter et al [35]							
Executive function/processing speed–focused CCT[r]	Use app 15 minutes/day 5 days/week	NR	NR	168.3 (69.0)	No	NR	NR
Verbal ability–focused CCT	Use app 15 minutes/day 5 days/week	NR	NR	363.8 (253.4)	No	NR	NR
O'Toole et al [36]							
LifeApp'tite	At discretion of therapists to decide frequency of app use	50 (83.3)	NR	NR	No	NR	NR
Place et al [37]							
Cogito	Record weekly audio notes on mood and complete weekly self-reports	NR	NR	NR	No	NR	NR
Proudfoot et al [40]							
MyCompass	Complete a minimum of 2 modules and monitor at least three moods or behaviors	NR	Mean 14.7 (SD 16.7)[p]	NR	No	NR	NR

Study and intervention name	App use instructions or adherence criteria	Rate of uptake[b], n (%)	Level of use		Duration of use		Completers[c], n (%) or %
			App uses	Minutes spent using the app, mean (SD)	Reported weekly use pattern	Used in the final week, n (%) or %	
Roepke et al [38]							
CBT-PPT SuperBetter	Use app 10 minutes/day	72 (77.4)	Mean 21.5 (SD 34.3), median 9.5[d,e]	NR	No	NR	31 (33.3)[s]
General SuperBetter	Use app 10 minutes/day	72 (74.23)	__j	NR	No	NR	64 (66.0)[s]
Stiles-Shields et al [39]							
Boost Me	No specific instructions reported	10 (100)[l]	Mean 97.7	NR	No	NR	NR
Thought Challenger	No specific instructions reported	7 (70)[l]	Mean 33.5	NR	No	NR	NR
Watts et al [13]							
Get Happy Program Mobile App	Complete 6 lessons and associated homework	15 (68.2)[l]	Mean 5.1 (SD 1.6)[e,t]	NR	No[u]	NR[u]	10 (45.5)[p]

[a]Table includes all treatment conditions that involved a mobile app component.

[b]Rate of uptake: number of participants randomized to the intervention who used it at least once.

[c]Completer: participants who completed the intervention as intended per intervention instructions or per specified adherence criteria.

[d]Reported metric cut across treatment groups.

[e]Only included participants who logged onto the app at least once.

[f]NR: not reported.

[g]Estimate based on figure, exact number not reported.

[h]Assumes participants who met adherence criteria during the last week also met adherence criteria in previous weeks. For example, in a 4-week intervention, those reported to have used the app in week 4 also used in weeks 1-3.

[i]Completion refers to meeting a specified adherence criteria involving app use not to complying with intervention use instructions.

[j]Metrics were only reported across conditions rather than for each group independently; all numbers are rounded to 1 decimal place.

[k]Owing to technical issues, data on rate of uptake were only available in 21 participants and data on ongoing use were only available in 18 participants. To calculate percentages presented, the number of people for whom data were available was used as the denominator.

[l]Based on reported numbers of participants who were randomized to the condition, but never started treatment. Reasons were not always related to willingness/interest in trying the relevant app. For example, reason may have been that the participant was unresponsive to outreach to inform them of their assigned treatment.

[m]Article reports "93.03% (SD 15.6) of patients randomized to the intervention group evaluated the subjective items in the MONARCA system on a daily basis." Unclear if this refers to participants using the system an average of 93.03% of days or if it refers to 93.03% of the participants in the intervention using it every day of the 6-month intervention period.

[n]As use data were self-reported, these metrics only include those participants who completed the posttreatment assessment. To calculate percentages presented, the total size of the treatment group was used as the denominator.

[o]Completion was defined by clinician contact not app use.

[p]Metric takes into account all participants randomized to the condition even if they did not log onto the app.

[q]Number represents the number of participants whose last use was week 7 or after.

[r]CCT: computerized cognitive training.

[s]Refers to the number of participants who downloaded all content.

[t]Uses refers to lessons completed.

[u]Number of lessons completed was reported, but lessons were not precisely 1 per week.

Discussion

Principal Findings

This scoping review has revealed that reporting on engagement with DMHIs in RCTs is highly variable. A number of basic metrics of intervention engagement, such as rate of intervention uptake, weekly use patterns, and number of intervention completers, were routinely not reported. When intervention engagement metrics were reported, it was common to see low levels of engagement. The variability in reporting and frequency of low engagement when reported highlight the importance of establishing minimum necessary reporting standards for engagement in DHMI research.

Only 64% (14/22) of studies included in this review specified rate of uptake, defined as the number of participants randomized to the intervention condition who used the app at least once. Past research suggests that rate of uptake cannot be assumed, especially in the context of fully remote, self-guided digital interventions. Those studies that did report this metric showed varied levels of uptake. For example, Arean at al [21] found that over one-half of participants did not download their assigned app, whereas Roepke et al [38] and Watts et al [13] found that closer to one-quarter of participants did not download their assigned app. The studies reviewed here varied in the type of app and design so different rates of uptake may be expected, but the extent of inconsistent reporting was surprising.

Level of use metrics, defined as both the number of app launches and the amount of time the intervention was used, was only reported in 59% (13/22) of the studies reviewed. These metrics—specifically, average number of uses and average time spent in the app—should be feasible to calculate when researchers have access to activity log data of the tested app, which was the case in most of the studies included. There can be some complications reporting these metrics. For example, it can be difficult to accurately report time spent in the app when participants leave an app open on their device for longer than they are actively using it. Similarly, apps can be launched only to be closed in a matter of seconds. However, in cases where these metrics are not appropriate for the intervention being evaluated, we would have expected to see alternative metrics such as number of clicks reported, but this was only the case in 1 of the reviewed studies [36].

Approximately one-quarter of studies (5/22, 23%) reported on participant duration of use, defined as reporting both weekly use patterns and the number of participants who used the app in the final week of the intervention period. It is well documented that, in general, mobile apps tend to be used heavily when first downloaded and that use decreases over time [41]. Similarly, concerns related to sustained engagement with web-based psychiatric interventions have been reported in routine-care implementation studies [4-6,17]. Inconsistent use of psychiatric intervention apps over time is an issue that needs to be addressed if our field is to mature; however, addressing this issue will be all the more difficult if such variations in use are not adequately reported in our published literature. Data from Dahne et al [25] provide an excellent example of how this metric is useful to report alongside level of use. They reported that 81.8% of participants in the intervention condition used the app at least eight times (an average of at least once per week), but only 51% of participants used the app during the last week. Much like patterns of use with other popular apps, these data suggest high initial use that declines over time.

In the context of intervention research, it is important to include some clear metric of intervention adherence or completion. Yet only 50% (11/22) of studies in this review clearly reported the number of participants considered to have completed the app-based components of the intervention as intended or other metrics related to completion such as percentage of patients who met a specified adherence threshold; percentage of patients completing clinician-based components of the intervention; and percentage of participants who downloaded all the intervention content. Just like psychotherapy or medication use, mobile app–based interventions incorporate some expected efficacious dose into the instructions for use. The fact that use can be accurately and objectively tracked from backend metrics is highly encouraging, and distinguishes our field from other treatment research (such as medication trials) where adherence has historically been extremely difficult to reliably measure. Further, completion need not be full use exactly as intended. For example, Arean et al [21] specified that 50% compliance with intervention instructions was considered completion. Simply not discussing who uses mobile app depression interventions as intended, however, will limit the potential for insight into and utility of these interventions.

Finally, one of our objectives in this review was to quantify standard level of engagement in RCTs of mobile app–based depression interventions. Our data extraction led us to conclude that with the current state of reporting, this is nearly impossible to do. What we did conclude is that engagement at all points—uptake, level, duration, and completion—is widely varied. Moreover, it was not uncommon to see completion rates at or below 50% of those participants randomized to a treatment condition (n=6) or to simply see engagement rates not reported at all (n=5).

Limitations

This scoping review has several limitations. First, this review illustrates an important dilemma in the field of DMHI research, but findings are limited to a subset of DMHI literature, specifically only that involving depression interventions in psychiatric samples with mobile app–based interventions. While we expect our proposed reporting guidelines to be useful across DMHIs, the extent to which the findings of this review carry through to mobile app interventions in other areas of mental health remains unclear. Second, our original goal in approaching this scoping review was to quantify typical engagement with DMHIs in RCTs; however, as we began the literature review, we ascertained that this goal would be difficult given the variability (and often absence) of metrics reported. This study, therefore, represents a shift in objectives. Third, we only reviewed papers from academic sources, which limits the kinds of mental health apps we took into account. The quality and objectivity of the data contained within independently published reports from private industries on their own mental health apps have yet to be reviewed. Finally, this review only evaluated literature though May 2020. While there is no reason to expect that reporting on engagement has improved, this work should be conceptualized as only a starting point for a discussion of appropriate reporting guidelines and future reviews or meta-analyses on this topic are warranted.

Conclusions

The emerging field of DMHIs has reached a critical juncture: intervention engagement has been widely recognized as the key factor limiting DMHI clinical utility. This review illustrates that engagement is variable and frequently underreported. Adopting a set of reporting guidelines that specify the minimum necessary information when publishing RCTs of DMHIs will provide new insights into how to improve engagement in mental health apps; allow for clear comparisons between DMHIs and other treatment

options; and offer benchmarks upon which further research must improve. Such reporting standards will complement the expanding literature on user-centered evaluations of engaging with digital health tools and interventions [42-44].

To this end, we suggest the 5-element framework applied in this study be used to guide minimum necessary DMHI engagement reporting standards. This framework includes the following: (1) intervention instructions or adherence criteria, defined as an explicit statement of what it means for participants to have used an intervention as intended or met some minimum intervention threshold; (2) rate of uptake, defined as the number of participants randomized to the intervention who downloaded the associated app(s) and used them at least once; (3) level of use metrics, defined as *both* the number of app launches and the amount of time the intervention was used (with alternative metrics such as number of clicks appropriate if more suitable for the intervention and justified); (4) duration of use, defined as participants' weekly use patterns; and (5) number of completers, defined as the number of participants who completed the intervention as intended per intervention instructions or per specified adherence criteria. We believe this framework could be a useful starting point to promote standards of reporting within the field, with room for future iterations.

Certainly complexities exist when identifying and reporting engagement with DMHIs given that these interventions vary widely in content and format. The reporting guidelines that we have suggested in response to our findings are intended both to be broadly applicable across DMHIs and to challenge the field to move past complexities and move toward greater transparency and rigor. We hope this begins an important discussion on reporting standards that will improve our understanding of how to evaluate and optimize DMHIs.

Acknowledgments

JML was partially supported by an NIMH Mentored Patient-Oriented Career Development Award (K23MH120324) and an NARSAD Young Investigator Grant from the Brian and Behavior Research Foundation. The authors acknowledge Britney Gluskin for her assistance with title and abstract screening.

Conflicts of Interest

JF is supported by a UK Research and Innovation Future Leaders Fellowship (MR/T021780/1) and has received honoraria / consultancy fees from Atheneum, Informa, Gillian Kenny Associates, Big Health, Nutritional Medicine Institute, ParachuteBH, Richmond Foundation and Nirakara, independent of this work.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) Checklist.
[DOCX File , 48 KB - jmir_v24i10e39204_app1.docx]

Multimedia Appendix 2
Other DMHI use data reported. DMHI: digital mental health intervention.
[DOCX File , 27 KB - jmir_v24i10e39204_app2.docx]

References

1. American Psychiatric Association. Practice Guideline for the Treatment of Patients With Major Depressive Disorder. Washington, D.C: American Psychiatric Association; 2010:1-152.
2. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
3. 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]
4. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, REEACT collaborative. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. Br J Psychiatry 2017 May;210(5):362-367. [doi: 10.1192/bjp.bp.116.192435] [Medline: 28254959]
5. Cavanagh K, Seccombe N, Lidbetter N. The Implementation of Computerized Cognitive Behavioural Therapies in a Service User-Led, Third Sector Self Help Clinic. Behav. Cogn. Psychother 2011 Feb 22;39(4):427-442. [doi: 10.1017/s1352465810000858]
6. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Cohen A, et al. A Web-Based Mental Health Platform for Individuals Seeking Specialized Mental Health Care Services: Multicenter Pragmatic Randomized Controlled Trial. J Med Internet Res 2019 Jun 04;21(6):e10838 [FREE Full text] [doi: 10.2196/10838] [Medline: 31165710]

7. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
8. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
9. Arnold C, Farhall J, Villagonzalo K, Sharma K, Thomas N. Engagement with online psychosocial interventions for psychosis: A review and synthesis of relevant factors. Internet Interv 2021 Sep;25:100411 [FREE Full text] [doi: 10.1016/j.invent.2021.100411] [Medline: 34401370]
10. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: Complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
11. 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]
12. 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]
13. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13(1):49. [doi: 10.1186/1471-244x-13-49]
14. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. J Med Internet Res 2017 Jan 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
15. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones – the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol. Med 2015 Jul 29;45(13):2691-2704. [doi: 10.1017/s0033291715000410]
16. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
17. 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]
18. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
19. 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]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
21. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
22. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
23. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker Mobile Web App With Email for Adults to Self-Manage Mild-to-Moderate Depression: Randomized Controlled Trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
24. 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):84726. [doi: 10.5812/ijpbs.84726]
25. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language Behavioral Activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
26. Dahne J, Lejuez C, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot Randomized Trial of a Self-Help Behavioral Activation Mobile App for Utilization in Primary Care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
27. 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]

28. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-Based Intervention (Run4Love) on Depressive Symptoms Among People Living With HIV in China: A Randomized Controlled Trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
29. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
30. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 09;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
31. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-Supported versus Full Behavioural Activation for Depression: A Randomised Controlled Trial. PLoS One 2015 May 26;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
32. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial. J Med Internet Res 2017 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
33. Moberg C, Niles A, Beermann D. Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
34. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, et al. Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. J Med Internet Res 2019 Aug 28;21(8):e13609 [FREE Full text] [doi: 10.2196/13609] [Medline: 31464192]
35. Motter JN, Grinberg A, Lieberman DH, Iqnaibi WB, Sneed JR. Computerized cognitive training in young adults with depressive symptoms: Effects on mood, cognition, and everyday functioning. J Affect Disord 2019 Feb 15;245:28-37. [doi: 10.1016/j.jad.2018.10.109] [Medline: 30366235]
36. O'Toole MS, Arendt MB, Pedersen CM. Testing an App-Assisted Treatment for Suicide Prevention in a Randomized Controlled Trial: Effects on Suicide Risk and Depression. Behav Ther 2019 Mar;50(2):421-429. [doi: 10.1016/j.beth.2018.07.007] [Medline: 30824256]
37. Place S, Blanch-Hartigan D, Smith V, Erb J, Marci CD, Ahern DK. Effect of a Mobile Monitoring System vs Usual Care on Depression Symptoms and Psychological Health: A Randomized Clinical Trial. JAMA Netw Open 2020 Jan 03;3(1):e1919403 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19403] [Medline: 31951270]
38. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
39. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: Pilot trial. Psychol Serv 2019 May;16(2):233-238 [FREE Full text] [doi: 10.1037/ser0000261] [Medline: 30407055]
40. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
41. Overall App Benchmarks H2 2017. Localytics. 2017. URL: https://uplandsoftware.com/localytics/resources/cheat-sheet/overall-app-benchmarks-h2-2017/#:~:text=User%20Retention,-Average%20Three%20Month&text=According%20to%20our%20data%2C%2043,month%20after%20they%20downloaded%20it [accessed 2022-09-01]
42. Nievas-Soriano BJ, García-Duarte S, Fernández-Alonso AM, Bonillo-Perales A, Parrón-Carreño T. Users evaluation of a Spanish eHealth pediatric website. Comput Methods Programs Biomed 2021 Nov;212:106462 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106462] [Medline: 34715515]
43. 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]
44. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]

Abbreviations

AMED: Allied and Complementary Medicine
DMHI: digital mental health intervention
HMIC: Health Management Information Consortium
HTA: Health Technology Assessment

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by T Leung; submitted 02.05.22; peer-reviewed by C Bedard, B Nievas Soriano; comments to author 29.05.22; revised version received 20.07.22; accepted 19.08.22; published 14.10.22.

Please cite as:
Lipschitz JM, Van Boxtel R, Torous J, Firth J, Lebovitz JG, Burdick KE, Hogan TP
Digital Mental Health Interventions for Depression: Scoping Review of User Engagement
J Med Internet Res 2022;24(10):e39204
URL: https://www.jmir.org/2022/10/e39204
doi:10.2196/39204
PMID:36240001

Review

The Effectiveness of Supervised Machine Learning in Screening and Diagnosing Voice Disorders: Systematic Review and Meta-analysis

Ghada Al-Hussain[1], MSc; Farag Shuweihdi[2], PhD; Haitham Alali[3], PhD; Mowafa Househ[4], PhD; Alaa Abd-alrazaq[5], PhD

[1]Department of Unified Health Record, Lean for Business Services, Riyadh, Saudi Arabia

[2]Leeds Institute of Health Sciences, School of Medicine, University of Leads, Leeds, United Kingdom

[3]Health Management Department, Faculty of Medical and Health Sciences, Liwa College of Technology, Abu Dhabi, United Arab Emirates

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

[5]AI Center for Precision Health, Weill Cornell Medicine, Doha, Qatar

Corresponding Author:
Alaa Abd-alrazaq, PhD
AI Center for Precision Health
Weill Cornell Medicine
Education City, Qatar Foundation
PO Box 24144
Doha
Qatar
Phone: 974 55708549
Email: alaa_alzoubi88@yahoo.com

Abstract

Background: When investigating voice disorders a series of processes are used when including voice screening and diagnosis. Both methods have limited standardized tests, which are affected by the clinician's experience and subjective judgment. Machine learning (ML) algorithms have been used as an objective tool in screening or diagnosing voice disorders. However, the effectiveness of ML algorithms in assessing and diagnosing voice disorders has not received sufficient scholarly attention.

Objective: This systematic review aimed to assess the effectiveness of ML algorithms in screening and diagnosing voice disorders.

Methods: An electronic search was conducted in 5 databases. Studies that examined the performance (accuracy, sensitivity, and specificity) of any ML algorithm in detecting pathological voice samples were included. Two reviewers independently selected the studies, extracted data from the included studies, and assessed the risk of bias. The methodological quality of each study was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 tool via RevMan 5 software (Cochrane Library). The characteristics of studies, population, and index tests were extracted, and meta-analyses were conducted to pool the accuracy, sensitivity, and specificity of ML techniques. The issue of heterogeneity was addressed by discussing possible sources and excluding studies when necessary.

Results: Of the 1409 records retrieved, 13 studies and 4079 participants were included in this review. A total of 13 ML techniques were used in the included studies, with the most common technique being least squares support vector machine. The pooled accuracy, sensitivity, and specificity of ML techniques in screening voice disorders were 93%, 96%, and 93%, respectively. Least squares support vector machine had the highest accuracy (99%), while the K-nearest neighbor algorithm had the highest sensitivity (98%) and specificity (98%). Quadric discriminant analysis achieved the lowest accuracy (91%), sensitivity (89%), and specificity (89%).

Conclusions: ML showed promising findings in the screening of voice disorders. However, the findings were not conclusive in diagnosing voice disorders owing to the limited number of studies that used ML for diagnostic purposes; thus, more investigations are needed. While it might not be possible to use ML alone as a substitute for current diagnostic tools, it may be used as a decision support tool for clinicians to assess their patients, which could improve the management process for assessment.

Trial Registration: PROSPERO CRD42020214438; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=214438

(J Med Internet Res 2022;24(10):e38472) doi:10.2196/38472

KEYWORDS

machine learning; voice disorders; systematic review; meta-analysis; diagnose; screening; mobile phone

Introduction

Background

Voice disorders are abnormalities in voice production that could be due to lesions or abnormal modifications in the structure of vocal folds [1]. In 2019, it was estimated that 16.9% of the population in Sweden had voice disorders [2], and in 2014, it was found that 1 in 13 adults in the United States develops voice disorders every year [3]. This led to a loss of US $845 million in the United States owing to missed working days among employees with voice disorders [4,5]. At the individual level, voice disorders can severely affect a patient's social life and mental health compared with other chronic disorders such as back pain [6]. Thus, 4.3% of the patients with voice disorders reported that they were unable to do certain job-related tasks due to the disorder [7]; this especially affects professions that have a high demand on the voice, for instance, teachers [8], singers, or telephone operators [9]. Therefore, screening or diagnosing voice disorders is essential to detect other related health conditions such as laryngeal lesions that could be a symptom of cancer [10]; thus, the diagnosis should be made as soon as possible [11,12].

Diagnosing and screening voice disorders involve auditory-perceptual and instrumental assessments. The auditory-perceptual assessment is carried out by a qualified speech and language therapist (SLT); in this assessment, the SLT determines the quality of patients' voice by listening to their sustained vowel production; for example, the, aa, or sound or continuous speech [13,14]. Furthermore, the instrumental assessment involves laryngeal imaging to examine the structure and function of vocal cords while the patient produces a vowel sound; other techniques are also used including video laryngeoendoscopy and video laryngostroboscopy examinations. In addition, acoustic instruments were used to analyze acoustic features (frequency, pitch, volume, and quality of sound) of voice samples of patients to assess voice disorders by using computer software [13-15]. Although the aforementioned assessments are recognized by the American Speech-Language-Hearing Association [13] and American Academy of Otolaryngology-Head and Neck Surgery [16], there is still a lack of standardized methods and guidelines to regulate these or other assessments [17]. Therefore, several limitations may pose a risk to the current assessment [18,19]. Although each case is evaluated objectively (via instrumental techniques, eg, stroboscopy), these objective tests include acoustic and visual imaging and videos; the acoustic techniques reveal the speech characteristics of the patient's speech sample, specifically, the frequency, intensity, loudness, and pitch, to give the clinician insight into other indicators such as the patient's rate of speech or voice; for example, the voice may be breathy or tremored [18]. Although these instrumental methods enable clinicians to perform objective tests, the validity of the tests largely depends on the auditory-perceptual skills of the clinician [18]. This is because the clinician first assesses the instrumental management or the patients' pathway and then chooses the type of instrumental assessment to be used. Naturally, any mistake in the auditory-perceptual assessment would affect the instrumental management, and thus, the whole management of the case; such subjective judgment might not be reliable as it relies on the clinician's skills and experience [18]. As the condition of each diagnosis or screening and the level of experience differ in each case, severe cases might be easier to diagnose or screen than mild cases; therefore, the experience of the SLT and the reliability of their judgment on each patient's condition differ, and low interrater correlations may occur (<0.9) [19]. Moreover, the agreement between experienced and inexperienced SLTs was found to be <75%, making the experience an essential part of the diagnosis or screening [20].

Machine learning (ML) was introduced for speech sounds in the early 1980s [21]. ML can be performed automatically by analyzing acoustic features either from voice recordings samples that are previously stored in a database such as the Massachusetts Eye and Ear Infirmary (MEEI), which are databases that stores a recordings of voice samples from patients in clinical environments, these recordings either recorded patients' voices while pronouncing vowels such as in MEEI [22] or continuous speech, or phrases such as in the Saarbruecken Voice Database [23]. ML is also used to analyze patients in the clinic by recording their voices via a microphone [1,21,24]. ML was applied either as a differential diagnosis for s, which involves diagnosing the voice sample as 1 of 2 diseases (voice disorders a or voice disorders b), or for screening different voice samples as either healthy or pathological voice. This method has been used to improve the diagnosis and screening process to be more objective. ML involves 2 different models: classification (supervised learning) and clustering or categorization (unsupervised learning) [25]. In the unsupervised model, the algorithm categorizes and identifies relationships within a data set [26]. By contrast, classification is a prediction model that defines labels, for example, disease or not disease, in clinical diagnosis [26], making it more common in diagnosing [27].

Research Problem and Aim

Although several studies have investigated the effectiveness of ML algorithms in detecting and diagnosing voice disorders, to the best of our knowledge, only 1 review attempted to summarize the evidence resulting from these studies [27]. However, there are several limitations in the review, including the following: it did not exclude studies that did not validate their ML outcomes by using validation techniques; it included studies that relied on scientific but not technical or objective solutions, and they relied on subjective assessment only; and it did not assess the included studies against any risk of bias assessment. Accordingly, this systematic review aimed to assess the effectiveness of supervised ML algorithms in screening and

diagnosing voice disorders. Thus, only supervised ML techniques were considered because supervised ML algorithms are more commonly used for diagnosing and detecting disorders.

Methods

This systematic review followed the Cochrane Library's systematic reviews for diagnostic test accuracy (DTA) guidelines [28] to meet the objectives of this review. The protocol for this review was registered with PROSPERO (CRD42020214438).

Search Strategy

Search Sources

The following 5 databases were searched on June 24, 2021: MEDLINE (via Ovid), Embase, Scopus, Web of Science, and ACM Digital Library. No language limitations were applied, and non-English articles were translated to check their applicability to the review. The retrieved references were exported and managed using EndNote 9.

Search Terms

A total of 2 groups of keywords were used to search the databases: one group representing the target diagnosis (ie, voice disorders) and the other group representing the intervention of interest (ie, ML algorithms). The terms were derived from ML and speech therapy experts. Medical Subject Headings were also included to maximize the sensitivity of the search in MEDLINE and Embase. The detailed search strategy that was applied to MEDLINE and Web of Science is shown in Multimedia Appendixes 1 and 2, respectively.

Eligibility Criteria

Inclusion Criteria

The population of interest in this review included patients diagnosed with a voice disorder. No restrictions were applied to the type of population characteristics (eg, age, gender, and ethnicity). With regard to index tests, we focused on supervised ML techniques (classification) that were used to screen or diagnose voice disorders in binary outcomes (eg, pathological voice vs healthy voice or voice disorder a vs b) by using voice samples collected in a controlled environment (eg, speech laboratories, hospitals, clinics, and databases). The reference standards of interest in this review are instrumental assessment and auditory-perceptual assessment, as both follow the recommendations of the American Speech-Language-Hearing Association [17] and American Academy of Otolaryngology [16]. To be included in this review, studies had to assess the diagnostic performance of ML algorithms by using at least one of the following outcomes: accuracy, sensitivity, and specificity. We included only peer-reviewed articles and empirical studies regardless of their study design. No restrictions were applied on the country of publication, year of publication, or language of publication.

Exclusion Criteria

We excluded studies that relied on clinicians' judgments only without using any instrumental tools to ensure the validity and reliability of the review, as relying on subjective assessment may be affected by the clinician's level of experience. Unsupervised ML methods were excluded. Conference papers, reviews, reports, editorials, ongoing studies, non–peer-reviewed articles, studies that assessed accuracy only, and those that did not assess sensitivity and specificity were also excluded.

Study Selection

Study selection was first conducted by screening the titles and abstracts of the retrieved studies. Although we excluded studies whose titles and abstracts did not meet any of the eligibility criteria, all studies that met the eligibility criteria or were unclear owing to a lack of information in their titles and abstracts were retained. We then read the full texts of the studies that remained after the title and abstract screening to assess their eligibility for this review. The study selection process was performed by 2 reviewers.

Data Extraction

The 2 reviewers created a data extraction form (Multimedia Appendix 3) and extracted the data from each included study. If a study did not report a required piece of information, we contacted the corresponding authors to obtain any missing information. If the corresponding authors did not reply within 2 weeks, we sent 2 reminders. If we did not receive a reply after 2 weeks of the second reminder, the missing piece of information was referred to as *n/a: not applicable* data were extracted in an Excel spreadsheet.

Evaluation of Methodological Quality

The risk of bias in the included studies was assessed using a revised tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 [29], which is highly recommended by the Cochrane Collaboration [30]. QUADAS-2 assessed the risk of bias in 4 domains in the included studies: patient selection, index test, reference standards, and flow and timing (Multimedia Appendices 4-7). Furthermore, QUADAS-2 appraised the applicability of the included studies to this review in terms of 3 domains: patient selection, index test, and reference standards. QUADAS-2 was modified to fit this review (Multimedia Appendix 8). The 2 reviewers assessed the methodological quality of all included studies by using Review Manager (RevMan version 5.4).

Data Synthesis and Analysis

Narrative and quantitative syntheses were conducted to analyze the outcome of each ML technique (accuracy, sensitivity, and specificity). If >1 study used the same ML technique, and the difference between the outcomes was not significant (<5%), the best outcome was considered in the meta-analysis. All outcomes are presented in the extraction table (Multimedia Appendix 3). In addition, if a study used voice samples from 2 different databases, each sample was included to account for the sample size (referred to as sample A and sample B in the forest plot).

The accuracy, sensitivity, and specificity of ML methods extracted from the eligible studies were analyzed using the random effect proportional meta-analysis to estimate a pooled proportion and 95% CI, which are based on the Wilson score [31] procedures. To stabilize the variances, the pooled estimate was calculated using the Freeman-Tukey double arcsine

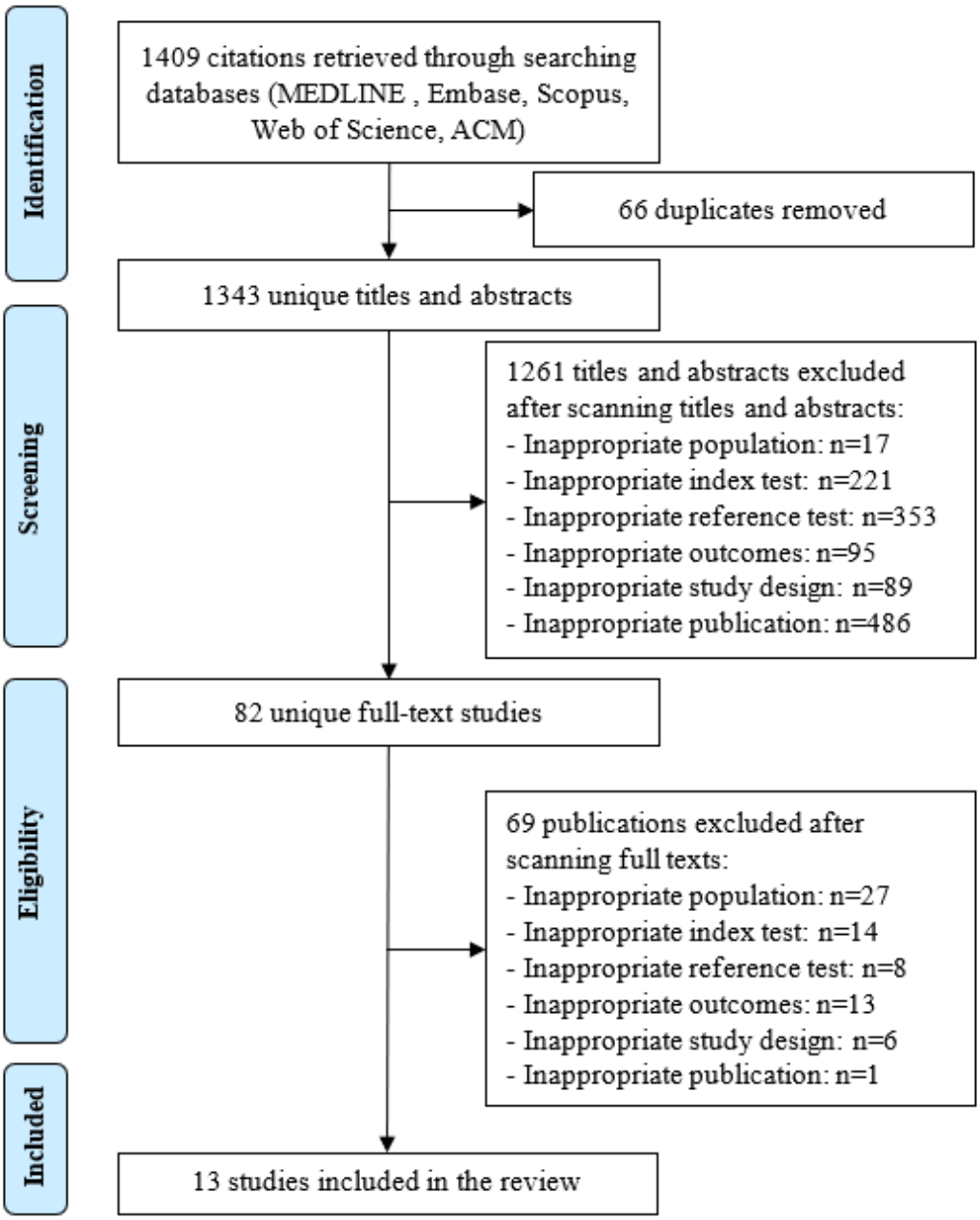
transformation [32], and heterogeneity was calculated using the I^2 measure [33]. A value of ≤50% is considered low, 51% to 75% moderate, and ≥76% high [33]. All results were plotted and presented in a forest plot. Studies were included in the meta-analysis if their scope of using ML was for screening. Statistical software STATA 16 was used to perform random effects meta-analyses.

Results

Search Results

As presented in Figure 1, a total of 1409 hits were identified by searching the 5 databases. No additional records were obtained from different resources. After removing duplicates, 95.31% (1343/1409) of articles were left. After scanning the titles and abstracts, 93.89% (1261/1343) of records were excluded, leaving 82 (6.11%) records for full-text reading. We excluded further 84% (69/82) of articles after full-text reading; therefore, only 16% (13/82) of studies were included in this review [34-46].

Figure 1. Flowchart of the study selection process.



Study Characteristics

Study Metadata

As shown in Table 1, the 13 included studies were conducted between 2000 and 2020. However, most of the studies (11/13, 85%) were conducted between 2010 and 2020. The year that witnessed the largest number of studies (3/13, 23%) was 2016. The included studies were conducted in 12 different countries, and approximately 30% (4/13) of them were conducted in Iran. All the studies were observational studies, peer-reviewed articles, and written in English.

Table 1. Metadata of the included studies.

Study	Year	Country	Publication language
Akbari and Arjmandi [34]	2015	Iran	English
Arias-Londoño et al [35]	2011	Greece	English
Arjmandi and Pooyan [36]	2012	Iran	English
Arjmandi et al [37]	2011	Iran	English
Cordeiro et al [38]	2017	Portugal	English
Ghasemzadeh et al [39]	2015	Iran	English
Godino-Llorente and Gómez-Vilda [40]	2004	Spain	English
Hadjitodorov et al [41]	2000	Bulgaria and France	English
Hariharan et al [42]	2014	Turkey	English
Lopes et al [43]	2017	Brazil	English
Mohmmad et al [44]	2020	Saudi Arabia and Malaysia	English
Souissi and Cherif [45]	2016	Tunis	English
Wang et al [46]	2011	China	English

Participants or Sample Characteristics

The number of participants or voice samples ranged from 40 to 960, with a total of 4019 and an average of 309 (Table 2). The included studies collected data from 6 different sources. The MEEI database was the most commonly used database among the included studies (9/13, 69%). Voice samples were collected from male and females and intersex participants in most included studies (12/13, 92%); however, 8% (1/13) of studies used voice samples from female participants only [43]. Participants' ages in the included studies ranged from 13 to 86 years, with an average age of 45 years (mean 46, SD 29.5 years).

Table 2. Characteristics of participants or sample.

Study	Voice sample size, n	Age (years), range	Male (%)	Setting or database	Database accessibility
Akbari and Arjmandi [34]	293	13-82	40	MEEI[a] database	Private
Arias-Londoño et al [35]	628	19-70	—[b]	MEEI and UPM[c] databases	Private
Arjmandi and Pooyan [36]	120	18-86	56	MEEI database	Private
Arjmandi et al [37]	100	16-85	67	MEEI database	Private
Cordeiro et al [38]	154	—	34	MEEI database	Private
Ghasemzadeh et al [39]	393	—	—	MEEI database	Private
Godino-Llorente and Gómez-Vilda [40]	135	—	—	MEEI database	Private
Hadjitodorov et al [41]	400	—	—	Phoniatric Department of the University Hospital in Sofia	Private
Hariharan et al [42]	274	20-68	—	MEEI and MAPACI databases	Private
Lopes et al [43]	279	18-65	0	Voice laboratory	Private
Mohmmad et al [44]	960	—	—	SVD[d]	Private
Souissi and Cherif [45]	120	—	—	SVD	Private
Wang et al [46]	226	26-58	—	MEEI database	Private

[a]MEEI: Massachusetts Eye and Ear Infirmary.

[b]Not available.

[c]UPM: Universidad Autónoma de Madrid.

[d]SVD: Saarbruecken Voice Database.

Index Test Characteristics

The included studies used 12 ML algorithms (Table 3). Least-squares support-vector machines (LS-SVMs) were the most used algorithms across studies (9/13, 69%), followed by quadratic discriminant analysis (QDA) (3/13, 23%) and K-nearest neighbor (K-NN) (4/13, 31%). The feature-extraction technique was reported in 85% (11/13) of studies. While 61% (8/13) of studies extracted short-term features (eg, mel frequency cepstral coefficients), 23% (3/13) extracted long-term features (eg, jitter and shimmer and fundamental frequency). A total of 3 feature reduction techniques were used in the included studies; linear discriminant analysis was the most used technique (4/13, 31%), and training-test split validation was the most prominent technique used in the included studies (10/13, 77%), followed by cross-validation technique (4/13, 31%).

Table 3. Index test characteristics.

Study	Machine learning method	Feature extraction	Feature reduction	Validation
Akbari and Arjmandi [34]	LS-SVM[a]	Mean, variance, skewness, kurtosis of coefficient, wavelet subband coefficients	Linear prediction analysis and LDA[b]	70% training and 30% 0% testing
Arias-Londoño et al [35]	LS-SVM	12 MFCC[c] and MSMR[d]	MSMR and LS-SVM	75% training and 25% testing (cross-validation–test split validation)
Arjmandi and Pooyan [36]	QDA[e], NMC[f], K-NN[g], LS-SVM, ML-NN[h], and PC[i]	PCA[j] and LDA; feature selection: IFS[k], FFS[l], BFS[m], and BBFS[n]	PCA and LDA	70% training and 30% validation
Arjmandi et al [37]	QDA, NMC, PC, K-NN, LS-SVM, and ML-NN	Fundamental frequency (average, high, and low variation), STD[o], PFR[p], jitter, shimmer, RAP[q], PPQ[r], smoothed PPQ, vAm[s], NHR[t], VTI[u], SPI[v], FTRI[w], ATRI[x], Tsam[y], T0[z], shimmer in dB, DVB[aa], DSH[ab], DUV[ac], NVB[ad], NSH[ae], and total number of segments pitch period during the period-to-period pitch extraction	PCA and LDA	70% training and 30% testing
Cordeiro et al [38]	SVM and DA[af]	MFCCs, line spectral frequencies, and delta-MFCC	N/A[ag]	75% training and 25% testing (k-fold cross-validation method, k=4; training-test split validation)
Ghasemzadeh et al [39]	ANN[ah] and LS-SVM	False neighbor fraction and mutual information	LDA and LS-SVM	70% training and 30% testing using cross-validation
Godino-Llorente and Gómez-Vilda [40]	LVQ[ai]	MFCC coefficient, energy, and first and second temporal derivatives	MFCC	70% training and 30% test split validation
Hadjitodorov et al [41]	K-NN	Pitch period (To), PPQ, APQ[aj], STAB[ak], the degree of the dissimilarity of the shape [47] of the pitch pulses, LHER[al], NHR, HNR[am], and energy in the pitch impulse-incepstra	LDA	Training-test split validation stage (200 phonation); testing (200 phonation)
Hariharan et al [42]	K-NN, LS-SVM, and GRNN[an]	5 level WPT[ao] decomposition	N/A	70% training and 30% testing using conventional validation and cross-validation
Lopes et al [43]	QDA	F0 measurements (mean and SD, jitter, shimmer, and GNE[ap])	N/A	Cross-validation
Mohmmad et al [44]	CNN[aq]	Octaves and its first and second derivatives	N/A	10-fold cross-validation
Souissi and Cherif [45]	LS-SVM and ANN	MFCC and first and second derivatives	MFCC, LDA, and delta	70% training; and 30% testing
Wang et al [46]	LS-SVM and GMM[ar]	36 dimensional MFCC parameters with 1 derivative were calculated every frame of 18-mel-cepstral coefficient	8, 16, and 32 mixture	10-fold cross-validation

[a]LS-SVM: least-squares support-vector machine.

[b]LDA: linear discriminant analysis.

[c]MFCC: mel frequency cepstral coefficient.

[d]MSMR: modulation spectra minimum redundancy.

[e]QDA: quadric discriminant analysis.

[f]NMC: neuromorphic computing.

[g]K-NN: K-nearest neighbor.

[h]ML-NN: multilayer neural network.

[i]PC: Parzen classifier.

[j]PCA: principal component analysis.

[k]IFS: individual feature selection.

[l]FFS: forward feature selection.

[m]BFS: backward feature selection.
[n]BBFS: branch-and-bound feature selection.
[o]STD: SD of fundamental frequency.
[p]PFR: phonatory fundamental frequency.
[q]RAP: relative average perturbation.
[r]PPQ: pitch perturbation quotient.
[s]vAm: peak amplitude variation.
[t]NHR: noise-to-harmonic ratio.
[u]VTI: voice turbulence index.
[v]SPI: soft phonation index.
[w]FTRI: Fo-tremor intensity index.
[x]ATRI: amplitude tremor intensity index.
[y]Tsam: length in seconds of analyzed voice data sample.
[z]T0: period of the average glottal period.
[aa]DVB: degree of voice breaks.
[ab]DSH: degree of subharmonic.
[ac]DUV: degree of voicelessness.
[ad]NVB: number of voice breaks.
[ae]NSH: number of subharmonic segments.
[af]DA: Discriminant analysis.
[ag]N/A: not applicable.
[ah]ANN: artificial neural network.
[ai]LVQ: learning vector quantization.
[aj]APQ: amplitude of the pitch pules.
[ak]STAB: stability of the t0 generation.
[al]LHER: low-high energy ratio.
[am]HNR: harmonics noise ratio.
[an]GRNN: general regression neural network.
[ao]WPT: wavelet packet transform.
[ap]GNE: glottal to noise excitation.
[aq]CNN: conventional neural network.
[ar]GMM: Gaussian mixture model.

Quality Assessment Results

Risk of Bias

In the patient selection domain, only 38% (5/13) of studies were judged to have a low risk of bias in patient sampling, as they used an appropriate sampling process to select voice samples (Multimedia Appendix 9). The risk of bias in index tests was rated as high in all included studies owing to the nature of the supervised ML tests, and their results were interpreted with prior knowledge of the results of the reference standard test. Owing to the subjective nature of voice assessment, it was not clear whether the reference standard correctly classified the patients. This led to an unclear risk of bias in the reference standard domain in all studies although the reference standard was used before the index test, and the findings were not affected by the findings of the index test. Patient flow and timing were poorly reported in almost all the studies (12/13, 92%). Thus, these studies were judged to pose an unclear risk of bias in terms of patient flow and timing. Multimedia Appendix 9 shows the QUADAS-2 tool risk of bias judgment in each included study across all 3 domains as well as applicability concerns for each study.

Applicability Concerns

There are no applicability concerns regarding how patients were selected in all included studies, as the patients' characteristics and the condition and setting of each test match the review question and criteria (Multimedia Appendix 9). Similarly, all included studies were judged to have low applicability concern in the index test as the ML algorithms method in the included studies matched the review definition of ML. However, the applicability concern in the reference standard was rated as unclear in 84% (11/13) of studies, as the voice samples in those studies were collected from databases, and the detailed diagnosis process of each voice sample was not described.

Performance of ML Algorithms

Diagnosing Voice Disorders

Only 8% (1/13) of studies used the QDA algorithm to differentiate between 2 [43]. As shown in Table 4, the accuracy, sensitivity, and specificity of the QDA ranged from 70% to 77%, 20% to 65%, and 74.76% to 95%, respectively. See the following section for a description of how QDA was used as a screening tool. For breakdown of the diagnostic findings, please refer to Multimedia Appendix 10.

Table 4. The performance of machine learning in diagnosing voice disorders.

Algorithm	Tested diseases	Accuracy (%)	Sensitivity (%)	Specificity (%)	Study
QDA[a]	Vocal polyps vs healthy	70.56	50	74.76	Lopes et al [43]
QDA	Vocal cyst vs healthy	72.67	60.83	78.1	Lopes et al [43]
QDA	Unilateral VF[b] paralysis or healthy	79.82	20	92.38	Lopes et al [43]
QDA	Middle-posterior triangular gap vs healthy	71.11	45	80.43	Lopes et al [43]
QDA	Sulcus vocalis vs healthy	78.75	50	83.33	Lopes et al [43]
QDA	VDDGER[c] vs healthy	72.44	33.33	90.71	Lopes et al [43]
QDA	Vocal nodules vs unilateral VF paralysis	76.61	20	88.57	Lopes et al [43]
QDA	Vocal nodules vs sulcus vocalis	72.68	50	75.95	Lopes et al [43]
QDA	Vocal nodules vs VDDGER	71	33.33	89.05	Lopes et al [43]
QDA	Vocal nodules vs sulcus vocalis	70	30	95	Lopes et al [43]
QDA	Vocal polyp vs healthy	75.14	65	78.33	Lopes et al [43]
QDA	Vocal cyst vs healthy	73.22	62.5	78.57	Lopes et al [43]

[a]QDA: quadratic discriminant analysis.

[b]VF: vocal fold.

[c]VDDGER: voice disorder due to gastroesophageal reflux.

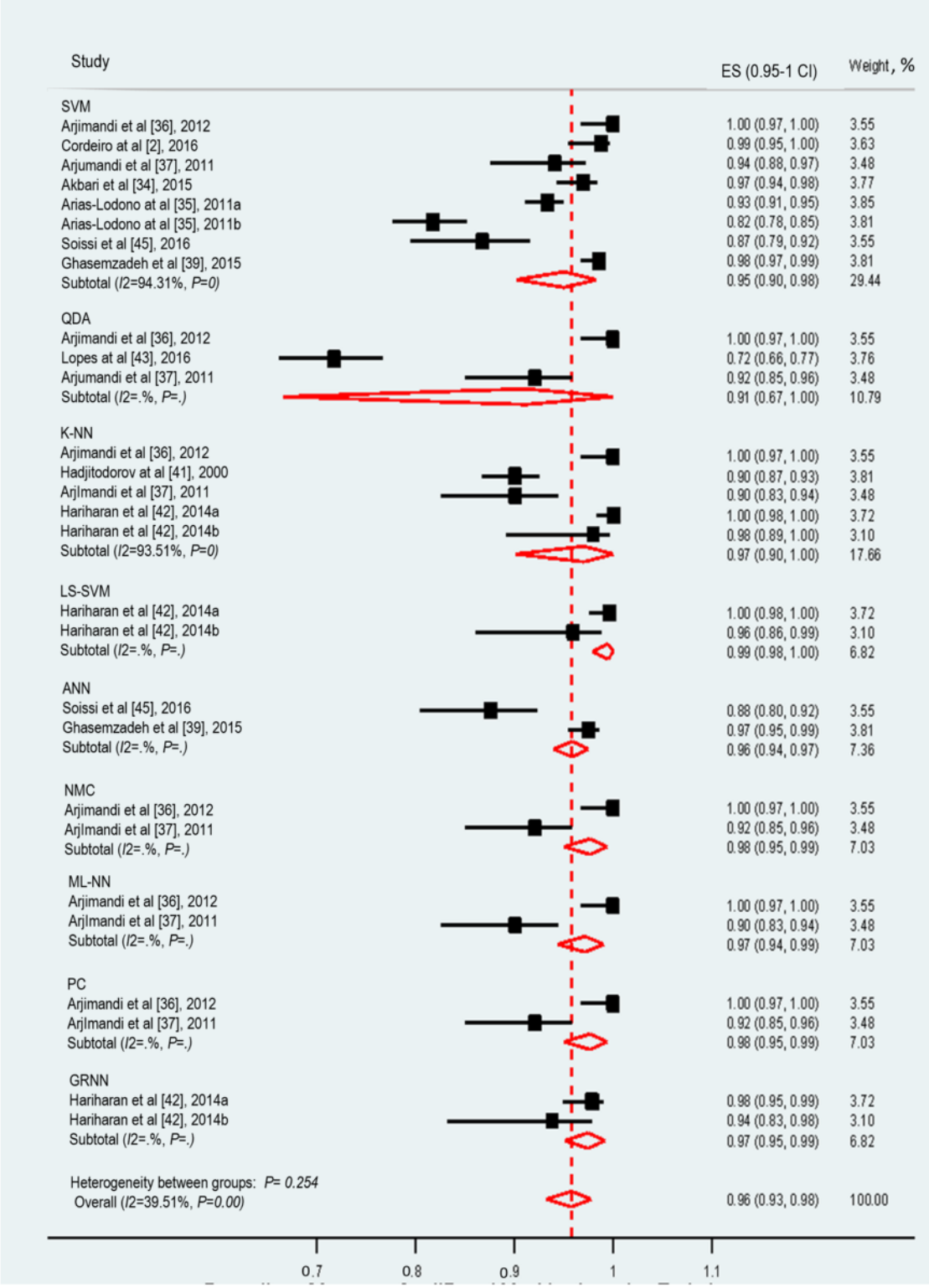
Screening Voice Disorders

Of the 13 included studies in the systematic review, 10 (77%) were included in the meta-analysis and 3 (23%) were excluded [39,40,44,46]. Of the 10 studies, 2 (20%) examined ML techniques by using 2 different databases: Arias-Londoño et al [35] (MEEI and Universidad Autónoma de Madrid [UPM] databases) and Hariharan et al [42] (MEEIEMPACI) Accordingly, the performance of ML techniques in these databases was included in the meta-analysis. More information about the performance in screening can be found in (Multimedia Appendix 11).

Accuracy

The accuracy of ML techniques in assessing voice disorders was reported in 77% (10/13) of studies. These studies examined the accuracy of 9 ML techniques. The pooled accuracy of the 9 ML techniques was 96% (95% CI 93%-98%; Figure 2). Significant heterogeneity was shown in the meta-analyzed studies (I^2=93.51%; P<.001), and the possible causes of this heterogeneity are discussed below. Regarding voice disorders assessment, the ML technique that achieved the highest accuracy was LS-SVM (99%), whereas the one that had the lowest accuracy was QDA (91%).

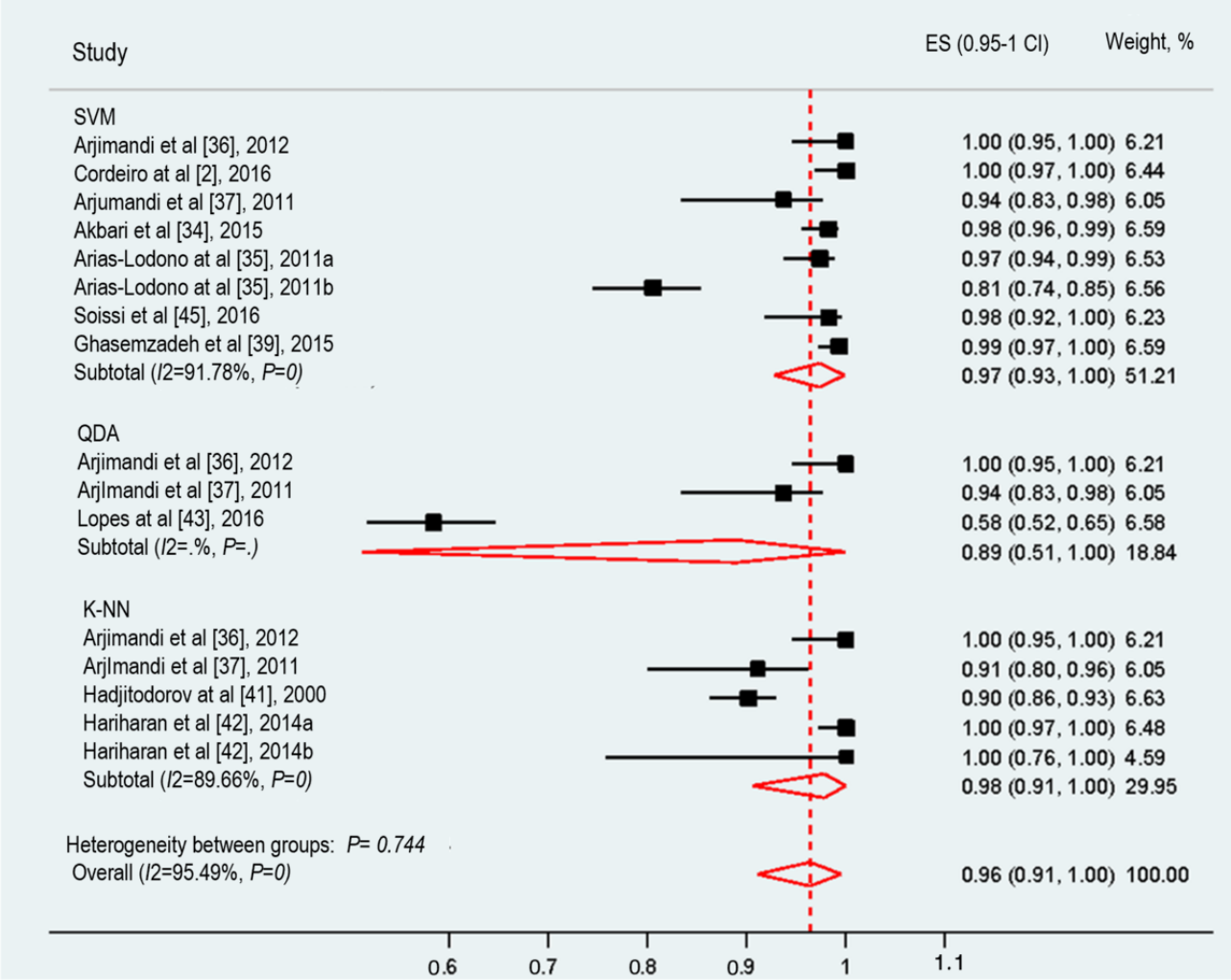
Figure 2. The forest plot shows the accuracy of machine learning algorithms in voice disorder screening. ANN: artificial neural network; GRNN: general regression neural network; K-NN: K-nearest neighbor; LS-SVM: least-squares support-vector machine; ML-NN: multilayer neural network; NMC: neuromorphic computing; PC: parzan Classifier; QDA: quadratic discriminant analysis; SVM: support vector machine.



Sensitivity

The sensitivity of ML techniques in assessing voice disorders was reported in 77% (10/13) of studies. These studies examined the sensitivity of 3 ML techniques. The pooled sensitivity of the 3 ML techniques was 96% (95% CI 91%-100%; Figure 3). The meta-analyzed studies showed significant heterogeneity (I^2=95.49%; P<.001), and the possible causes of such heterogeneity are discussed in further sections. K-NN had the highest sensitivity (98%) among the 3 ML techniques, while QDA achieved the lowest sensitivity (89%).

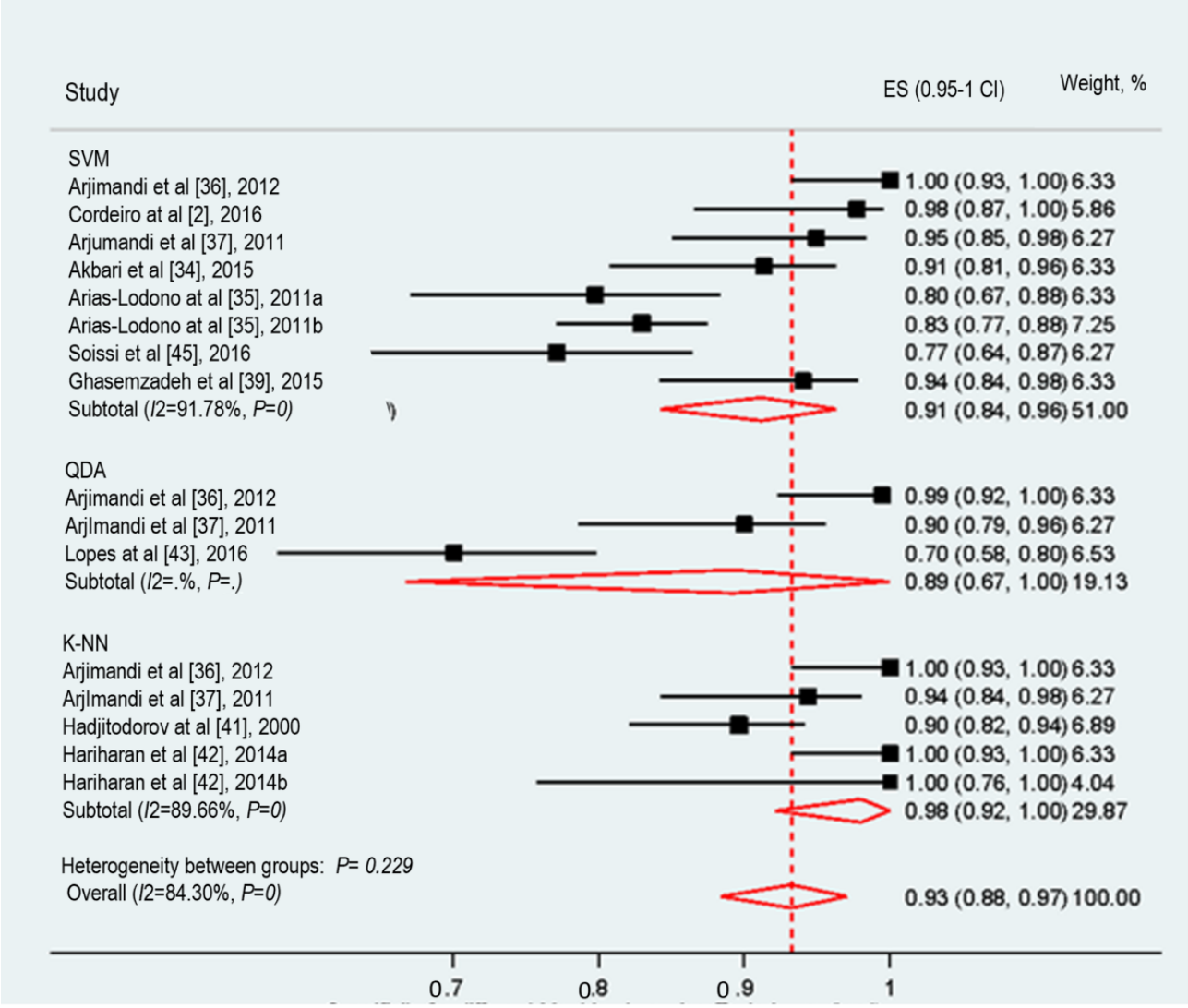
Figure 3. The forest plot shows the sensitivity of machine learning algorithms in voice disorder screening. K-NN: K-nearest neighbor; QDA: quadratic discriminant analysis; SVM: support vector machine.



Specificity

The specificity of ML techniques in assessing voice disorders was examined in 77% (10/13) of studies and included the specificity of 3 ML techniques. The pooled specificity of the 3 ML techniques was 93% (95% CI 88%-97%; Figure 4). The meta-analyzed evidence showed significant heterogeneity (I^2=84.3%; P<.001); the possible causes of heterogeneity are discussed below. The ML technique that achieved the highest specificity was K-NN (98%), whereas the one that had the lowest specificity was QDA (89%).

Figure 4. The forest plot shows the specificity of machine learning algorithms in voice disorder screening. K-NN: K-nearest neighbor; QDA: quadratic discriminant analysis; SVM: support vector machine.



Heterogeneity and Pooled Performance

The possible source of heterogeneity in the pooled performance was explored, and the possibility that studies that used short-term parameters, such as the study by Arjmandi et al [37], increased the heterogeneity in K-NN and LS-SVM was found. In the K-NN algorithm, the heterogeneity was reduced to 69.73% when the study by Arjmandi et al [37] (which used long-term parameters) was excluded; in specificity, it was 84.92% in sensitivity and 91.55% in accuracy. This was also found in the study by Hadjitodorov et al [41], which also used long-term parameters, and when it was excluded, the heterogeneity in all K-NN outcomes was reduced (Multimedia Appendix 12 presents further details on the heterogeneity values when each study was removed). Similarly, when the study by Arjmandi et al [37] was removed from the LS-SVM forest plot for sensitivity, a reduction was found in I^2 test values, which decreased to 91.89%. Therefore, long-term parameters could affect the sensitivity of LS-SVM and all 3 outcomes in K-NN. Furthermore, the database used by Arias-Londoño et al [35] and Souissi and Cherif [45] might increase the heterogeneity in LS-SVM performance. Arias-Londoño et al [35] used the UPM database, which is a Spanish sounds database, thus excluding the study from the sensitivity and specificity forest plot of LS-SVM, which decreased the heterogeneity to 58% and 71%, respectively. On the other hand, Souissi and Cherif [45] used voice samples from S (German speech samples database), whereas the remaining studies used s from the English speech samples database (MEEI; Multimedia Appendix 12).

Discussion

Principal Findings

This study systematically reviewed the performance of ML in assessing voice disorders, similar to another study by Syed et al [27] that examined the accuracy of ML algorithms at the voice database level and qualitatively analyzed the accuracy of each ML algorithm technique. It was concluded that LS-SVM is the most common algorithm used in studies included in this research, which aligns with our findings. Furthermore, the performance showed the accuracy of LS-SVM to be >93%, which was similar to our findings. Generally, ML performance was found to be more promising when it was used as a screening tool rather than in diagnosis, achieving >90% in all 3 outcomes

(accuracy, sensitivity, and specificity). Second, the findings differ significantly between the algorithms or even within the same algorithm in different studies. For example, LS-SVM was almost 100% in all 3 outcomes; however, Parzen classifier showed sensitivity ranging from 74% to 100%. Because of the limited number of studies, the performance of ML in ≤2 studies remains unclear. This was also noticed in ML algorithms that were used in the diagnosis, as only 1 study implemented ML algorithms to differentiate between different disorders (diagnosis). For example, the performance of QDA in screening showed 83% accuracy, 91% sensitivity, and 68% specificity. By contrast, it was found to be <76% in diagnosis, and the percentage fell sharply in sensitivity and specificity in the same study [43]. However, this finding could not be conclusive because of the limited number of studies that used ML for diagnosis (1 study).

The analysis implies that K-NN and LS-SVM showed the highest accuracy. K-NN demonstrated increased specificity; however, LS-SVM was found to be better at detecting true positive cases. Because ML in the included studies was used as a screening tool (pathological voice vs healthy voice), the ability of ML to be more sensitive might be more important than the ability to be specific. This may be due to the consequences of diagnosing healthy voiced patients as pathological voice which will only lead to further examination (stroboscopy). Moreover, it will not cause any distress to the patient, as the diagnosis is not final, and patients would only be referred for further examination. However, in less sensitive tests, misdiagnosis of patients can lead to harmful consequences.

Research and Practical Implications

Practical Implications

When a person's strength, agility, and structure of vocal folds result in pathological noise and reduced acoustic tone, their vocal pathology may be serious enough to qualify as a voice disorder. These disorders can be caused by tissue diseases and changes in tissue, mechanical stress, surface discomfort, systemic changes, changes in muscles and nerves, and many other factors [48]. Research on has achieved a wide scope, partly because of its societal benefits. Standard databases have been developed to mitigate disorders and include new features and emphasis on specific voice disorders while using deep neural networks. Recently, subjective and objective evaluations of vocal issues have received considerable attention in the research field [49].

Subjective assessments may be conducted by clinicians, as they focus on the patient's voice and use different instruments to discern various vocal disorder diagnoses. ML can be used as a decision-support tool for clinicians conducting auditory-perceptual assessments [14]. A second assessment, known as "target evaluated assessments," focuses on the automatic, computer-based processing of acoustic signals. These signals assess and recognize the underlying vocal pathology, which may not be screened or diagnosed by a clinician [50]. Consequently, this type of evaluation is nonsubjective. Furthermore, when using this type of assessment, voices can be captured and stored at a global level via cloud technologies by using various intelligent devices. This has been beneficial for researchers across the globe, who can access the data through different academic institutions.

Using ML as an assessment tool may reduce the learning gap between experienced and inexperienced clinicians. Bassich and Ludlow [20] found that the intrajudge test-retest agreement was <75% when evaluating voice quality in patients with polyps or vocal fold nodules; thus, the overall reliance on experienced clinicians in voice assessment might be eliminated. Furthermore, the practice of using instrumental assessments in practice could be eliminated, as ML may reduce the need to conduct instrumental assessments for more typical cases [27]. However, eliminating instrumental assessments altogether may lead to misdiagnoses, for example, if a patient with laryngeal cancer was screened "as healthy," the clinician may not have performed a stroboscopic examination. Therefore, we aim to further our study by establishing an ideal and automatic ML-based system. We anticipate that this system will be sensitive, accurate, efficient, and successful in detecting and diagnosing various voice disorders quickly and effortlessly for both patients and practitioners.

The review showed that ML provided optimum performance in screening and diagnosing voice disorders to inform clinicians of anomalies. A comparison of the performance of ML algorithms, including accuracy, specificity, and sensitivity, across studies is recommended owing to the different characteristics of each study. The most commonly used ML methods for diagnosing voice disorders in this review were LS-SVM and artificial neural network algorithms. However, the preference of applying 1 ML method to another was not clearly explained in the studies. All studies used internal validation (training and test splits and cross-validation) to evaluate the ML quality. However, external validation is a necessary procedure to evaluate the real quality of ML predictions for new data. Therefore, external validation is essential to implement ML in routine clinical practice to diagnose voice disorders. Therefore, external validation must be performed before using ML for any clinical diagnosis. None of the ML methods investigated in this review used external validation.

Implications for Research

This paper analyzes the literature related to the effectiveness of using ML algorithms to screen and diagnose voice disorders. It not only provides insight into the type of research conducted over the last 2 decades but also highlights the areas of research needing further experimentation and analysis. Researchers and practitioners can use this research to improve their objective screening or diagnosis of speech pathology. For instances, voice disorders [23], MEEI [22], and UPM databases [51] are all accessible to researchers interested in voice disorders case studies. However, these data repositories are not without their flaws. For instance, certain databases are uniformly classified into healthy and unhealthy classes. These voices are, in turn, generally categorized as "healthy" and "pathological" in most of the research published using these data. Some databases do not specify the severity of voice disorders or provide sufficient details on the pathological symptoms during phonation. As such, some samples may appear healthy normal despite being labeled

as pathological, and vice versa. In addition, >1 disorder may be used to label documents, which can be challenging to incorporate or exclude samples in different languages [52]. The nature of supervised ML, that is, "labeled," tests require prior knowledge of the reference standard finding to the corresponding test. This may lead to a higher risk of bias in some quality assessment tools, such as the QUADAS-2 tool, which shows a high risk of bias in the index test domain. Future researchers may wish to consider providing information on how a reference standard was applied when examining the performance of ML. Furthermore, these repositories may determine a more specific judgment on suitable demographic characteristics and how to appropriately classify these specifics. Finally, differential diagnostic abilities for ML may be better examined by dividing both the outcomes of each disorder as well as their severity. This would allow for more definitive and specific findings about the type of patients for whom ML may be more effectively used.

Because ML in the included studies was used as a screening tool (pathological voice vs healthy voice), the ability of ML to be more sensitive might be more important than its ability to be more specific. This may be due to the consequences of diagnosing healthy patients as unhealthy (patients with pathological voice), which will lead to further examination (stroboscopy) and not cause patient distress, as the diagnosis, at this point, is not final and patients would be referred for further examinations. Misdiagnosing patients (less sensitive tests) could lead to harmful consequences and distress, for example, if life-threatening diseases such as laryngeal cancer are misdiagnosed.

It should also be considered that ML can be used as a decision-support tool by clinicians while subjectively judging patients' voices to determine whether they should undergo further examinations. Applying the ML algorithm as a screening tool could help in predetermining the patient's voice condition. Consequently, this could support the clinicians' whole management process in voice disorders assessment, especially in their decision on whether to apply an instrumental examination for the patient, a decision that is currently being made subjectively. Therefore, applying ML as a screening tool would reduce the gap between experienced and inexperienced clinicians (the agreement was found to be <75%) [20], and the overall reliance on experienced clinicians in voice assessment might be eliminated. Furthermore, the use of instrumental assessments in practice could be eliminated, as not all patients will have to undergo instrumental assessments (ML might reduce the need to use them for healthy cases). Therefore, the cost of assessing voice disorders might be reduced.

Our findings also imply that ML can be used in web-based methods to detect voice disorders. This means that the algorithms can be used in smartphone apps or users' phone calls to detect the presence of voice disorders or even track the progress of their therapy. This might eliminate the amount of time spent by the clinician to screen or diagnose or record the progress of each follow-up. This study also found that researchers may want to consider investigating the applicability of various ML algorithms to identify and diagnose voice disorders. moreover, adding to previously established databases is recommended, which includes adding different languages, such as the Arabic voice pathology database, to other mainstream repositories.

Strengths

The key strength of this review is that it follows the DTA systematic review and search strategy. First, this review was in accordance with the Cochrane Library DTA systematic reviews, and second, it used a variety of medical, computer, and engineering databases. This increased the sensitivity of the review and broadened the search, overcoming the limited number of related articles. Moreover, in the screening process, in cases where the relevance of the abstract was not clear, the study was included in the full-text scanning. This eliminated any chance of eliminating relevant articles from the review. In addition, in the reference standard test, the inclusion criteria were restricted to a controlled environment, which might have ensured a more accurate and reliable result.

This is the first review to systematically assess the performance of different ML algorithms in the assessment and diagnosis of voice disorders. A total of 13 observational studies were included, which recruited patients from both genders and different age groups (13-85 years). In all, 14 ML techniques were tested, 9 of which were included in the meta-analysis, and their pooled accuracies, sensitivities, and specificities were estimated.

Limitations

The main weakness of this review is the limited reporting by primary studies; for example, the criteria for selecting voice samples from the databases or the patient recruitment process, the poor reporting of the demographic characteristics of the sample, and the severity of the voice disorders in each case. This hindered the ability to find sources of heterogeneity, as subgroup analysis based on gender, age group or type, or severity of each disease could not be investigated. Furthermore, the main outcomes of the review could not be more specific to a certain gender or age group or the type or severity of the disease. Mentioning these details could have allowed for further investigation of which factors—voice disorders, gender, or age group—would determine the accuracy of ML performance. In the patient selection domain, more than half (8/13, 60%) of the included studies demonstrated an unclear risk of bias. The poor reporting of how voice samples were chosen from the database led to the estimated accuracy being subject to bias. The bias increased when the voice samples were not chosen randomly, as they might have been chosen based on unreported severities. However, removing these studies from the meta-analysis was not possible owing to the limited number of included studies.

All included studies (13/13, 100%) failed to report how the reference standard was used, thus leading to an "unclear" risk of bias assessment in the overall reference standard. This is mainly due to the use of voice samples from a database; therefore, the clinicians' assessment was not performed by the authors of the primary studies. Moreover, the clinicians' assessment, which was applied by the chosen database, was not reported in the studies. Not knowing how the assessment was performed increased the risk of bias, and the outcome of the

review was found to be unclear. Although the authors were contacted to request further details about the choice of voice samples and reference standard assessment, no response was received. Poor reporting led to an unclear risk of bias in the flow and timing of patients in almost all included studies (12/13, 92%), especially the lack of reporting of the time intervals between clinicians' assessment and the recording of patients' voices. For example, if the recordings were made at intervals of a few months after the clinician's assessment, the patients' condition could have changed from when the first recording was made. Consequently, this increased the chance of misclassification or misdiagnosis, as the voice sample diagnosis could be different from the clinician's diagnosis. Better reporting of patients' diagnosis and recruitment process would lead to a clearer risk of bias assessment.

Conclusions

ML showed promising findings in screening, as its accuracy, sensitivity, and specificity showed high performance. The findings also suggested that ML can be further used in new smartphone apps for screening purposes and that screening can be conducted on the web. In scholarly research, more research with specific patient demographics and disorders is recommended. However, definitive conclusions could not be drawn about the effectiveness of ML in diagnosing owing to the limited number of studies (only 1). Therefore, we recommend using ML as a decision-support tool for clinicians during screening. For more definitive conclusions regarding the use of ML in diagnosis, more studies are suggested to be conducted, and risk of bias assessment that suits the application of ML for medical purposes and supervised ML is encouraged.

Acknowledgments

The authors would like to express their sincere gratitude to Munerah Al-Abdulsalam, biomedical engineer, for her support and guidance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy in Web of Science.
[PNG File , 196 KB - jmir_v24i10e38472_app1.png]

Multimedia Appendix 2
Search strategy in MEDLINE.
[PNG File , 117 KB - jmir_v24i10e38472_app2.png]

Multimedia Appendix 3
Extraction table.
[DOCX File , 14 KB - jmir_v24i10e38472_app3.docx]

Multimedia Appendix 4
Quality Assessment of Diagnostic Accuracy Studies 2 tool; patient selection domain.
[PNG File , 97 KB - jmir_v24i10e38472_app4.png]

Multimedia Appendix 5
Quality Assessment of Diagnostic Accuracy Studies 2 tool; index test domain.
[PNG File , 99 KB - jmir_v24i10e38472_app5.png]

Multimedia Appendix 6
Quality Assessment of Diagnostic Accuracy Studies 2 tool; reference standard domain.
[PNG File , 101 KB - jmir_v24i10e38472_app6.png]

Multimedia Appendix 7
Quality Assessment of Diagnostic Accuracy Studies 2 tool; flow and timing domain.
[PNG File , 102 KB - jmir_v24i10e38472_app7.png]

Multimedia Appendix 8
Modified Quality Assessment of Diagnostic Accuracy Studies 2 tool for this systematic review.
[PNG File , 200 KB - jmir_v24i10e38472_app8.png]

Multimedia Appendix 9
Quality Assessment of Diagnostic Accuracy Studies 2 tool risk of bias judgment in each included study across all domains and applicability concerns.
[PNG File , 96 KB - jmir_v24i10e38472_app9.png]

Multimedia Appendix 10
Summary of the performance of machine learning algorithms that were used for diagnosis.
[PNG File , 40 KB - jmir_v24i10e38472_app10.png]

Multimedia Appendix 11
Summary of the performance of machine learning algorithms that were used in screening.
[PNG File , 39 KB - jmir_v24i10e38472_app11.png]

Multimedia Appendix 12
The heterogeneity values after removing heterogenous studies from each forest plot.
[PNG File , 62 KB - jmir_v24i10e38472_app12.png]

References

1. Baken RJ, Orlikoff RF. Clinical Measurement of Speech and Voice. 2nd edition. Baltimore, MD, USA: Singular Publishing; 2000.
2. Lyberg-Åhlander V, Rydell R, Fredlund P, Magnusson C, Wilén S. Prevalence of voice disorders in the general population, based on the Stockholm public health cohort. J Voice 2019 Nov;33(6):900-905. [doi: 10.1016/j.jvoice.2018.07.007] [Medline: 30126692]
3. Bhattacharyya N. The prevalence of voice problems among adults in the United States. Laryngoscope 2014 Oct;124(10):2359-2362. [doi: 10.1002/lary.24740] [Medline: 24782443]
4. Zabret M, Hočevar Boltežar I, Šereg Bahar M. The importance of the occupational vocal load for the occurence and treatment of organic voice disorders. Zdr Varst 2018 Mar;57(1):17-24 [FREE Full text] [doi: 10.2478/sjph-2018-0003] [Medline: 29651311]
5. Cohen SM, Kim J, Roy N, Asche C, Courey M. The impact of laryngeal disorders on work-related dysfunction. Laryngoscope 2012 Jul;122(7):1589-1594. [doi: 10.1002/lary.23197] [Medline: 22549455]
6. Benninger MS, Ahuja AS, Gardner G, Grywalski C. Assessing outcomes for dysphonic patients. J Voice 1998 Dec;12(4):540-550. [doi: 10.1016/s0892-1997(98)80063-5] [Medline: 9988041]
7. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. Laryngoscope 2005 Nov;115(11):1988-1995. [doi: 10.1097/01.mlg.0000179174.32345.41] [Medline: 16319611]
8. Vilkman E. Voice problems at work: a challenge for occupational safety and health arrangement. Folia Phoniatr Logop 2000;52(1-3):120-125. [doi: 10.1159/000021519] [Medline: 10474011]
9. Koufman JA, Isaacson G. The spectrum of vocal dysfunction. Otolaryngol Clin North Am 1991 Oct;24(5):985-988. [Medline: 1754226]
10. Resouly A, Hope A, Thomas S. A rapid access husky voice clinic: useful in diagnosing laryngeal pathology. J Laryngol Otol 2001 Dec;115(12):978-980. [doi: 10.1258/0022215011909800] [Medline: 11779327]
11. Dankovičová Z, Sovák D, Drotár P, Vokorokos L. Machine learning approach to dysphonia detection. Appl Sci 2018 Oct 15;8(10):1927. [doi: 10.3390/app8101927]
12. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 2011 Jun 06;8(59):842-855 [FREE Full text] [doi: 10.1098/rsif.2010.0456] [Medline: 21084338]
13. Autism Spectrum Disorder. American Speech-Language-Hearing Association. 2022. URL: https://www.asha.org/practice-portal/clinical-topics/autism/ [accessed 2022-03-08]
14. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. Am J Speech Lang Pathol 2009 May;18(2):124-132. [doi: 10.1044/1058-0360(2008/08-0017)] [Medline: 18930908]
15. Fawcus M. Hyperfunctional voice: the misuse and abuse syndrome. In: Fawcus M, editor. Voice Disorders and their Management. New York, NY, USA: Springer; 1991:139-175.
16. Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ, et al. Clinical practice guideline: hoarseness (dysphonia) (update) executive summary. Otolaryngol Head Neck Surg 2018 Mar;158(3):409-426. [doi: 10.1177/0194599817751031] [Medline: 29494316]
17. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental

assessment of vocal function. Am J Speech Lang Pathol 2018 Aug 06;27(3):887-905. [doi: 10.1044/2018_AJSLP-17-0009] [Medline: 29955816]
18. Duffy J. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. Boston, MA, USA: Springer; 2019.
19. Maier A, Haderlein T, Eysholdt U, Rosanowski F, Batliner A, Schuster M, et al. PEAKS – a system for the automatic evaluation of voice and speech disorders. Speech Commun 2009 May;51(5):425-437. [doi: 10.1016/j.specom.2009.01.004]
20. Bassich CJ, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. J Speech Hear Disord 1986 May;51(2):125-133. [doi: 10.1044/jshd.5102.125] [Medline: 3702360]
21. Hegde S, Shetty S, Rai S, Dodderi T. A survey on machine learning approaches for automatic detection of voice disorders. J Voice 2019 Nov;33(6):947.e11-947.e33. [doi: 10.1016/j.jvoice.2018.07.014] [Medline: 30316551]
22. Lab Database | Laboratory on the Physiology, Acoustics, and Perception of Speech. MIT OpenCourseWare. 2022. URL: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-542j-laboratory-on-the-physiology-acoustics-and-perception-of-speech-fall-2005/lab-database/ [accessed 2022-06-12]
23. Pützer M, Barry W. Saarbrücken Voice Database-Handbook. Deutsche Forschungsgemeinschaft. 2022. URL: http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4 [accessed 2022-06-12]
24. Jackson-Menaldi MC. La voz patológica. Madrid, Spain: Editorial Médica Panamericana; 2002.
25. Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, et al. Machine learning in bioinformatics. Brief Bioinform 2006 Mar;7(1):86-112. [doi: 10.1093/bib/bbk007] [Medline: 16761367]
26. Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clin Epigenetics 2020 Apr 03;12(1):51 [FREE Full text] [doi: 10.1186/s13148-020-00842-4] [Medline: 32245523]
27. Syed SA, Rashid M, Hussain S. Meta-analysis of voice disorders databases and applied machine learning techniques. Math Biosci Eng 2020 Nov 11;17(6):7958-7979 [FREE Full text] [doi: 10.3934/mbe.2020404] [Medline: 33378928]
28. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy. Cochrane Methods. 2010. URL: https://methods.cochrane.org/sdt/ [accessed 2022-06-12]
29. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
30. Deeks JJ, Bossuyt PM, Gatsonis C. Cochrane handbook for systematic reviews of diagnostic test accuracy. The Cochrane Collaboration. 2009. URL: https://methods.cochrane.org/sdt/ [accessed 2022-06-12]
31. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998 Apr 30;17(8):857-872. [doi: 10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e] [Medline: 9595616]
32. Freeman MF, Tukey JW. Transformations related to the angular and the square root. Ann Math Statist 1950 Dec;21(4):607-611. [doi: 10.1214/aoms/1177729756]
33. Higgins JP, 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]
34. Akbari A, Arjmandi MK. Employing linear prediction residual signal of wavelet sub-bands in automatic detection of laryngeal pathology. Biomed Signal Process Control 2015 Apr;18:293-302. [doi: 10.1016/j.bspc.2015.02.008]
35. Arias-Londoño JD, Godino-Llorente JI, Markaki M, Stylianou Y. On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices. Logoped Phoniatr Vocol 2011 Jul;36(2):60-69. [doi: 10.3109/14015439.2010.528788] [Medline: 21073260]
36. Arjmandi MK, Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. Biomed Signal Process Control 2012 Jan;7(1):3-19. [doi: 10.1016/j.bspc.2011.03.010]
37. Arjmandi MK, Pooyan M, Mikaili M, Vali M, Moqarehzadeh A. Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. J Voice 2011 Nov;25(6):e275-e289. [doi: 10.1016/j.jvoice.2010.08.003] [Medline: 21186096]
38. Cordeiro H, Fonseca J, Guimarães I, Meneses C. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. J Voice 2017 May;31(3):384.e9-384.14. [doi: 10.1016/j.jvoice.2016.09.003] [Medline: 27743845]
39. Ghasemzadeh H, Tajik Khass M, Khalil Arjmandi M, Pooyan M. Detection of vocal disorders based on phase space parameters and Lyapunov spectrum. Biomed Signal Process Control 2015 Sep;22:135-145. [doi: 10.1016/j.bspc.2015.07.002]
40. Godino-Llorente JI, Gómez-Vilda P. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. IEEE Trans Biomed Eng 2004 Feb;51(2):380-384. [doi: 10.1109/TBME.2003.820386] [Medline: 14765711]
41. Hadjitodorov S, Boyanov B, Teston B. Laryngeal pathology detection by means of class-specific neural maps. IEEE Trans Inf Technol Biomed 2000 Mar;4(1):68-73. [doi: 10.1109/4233.826861] [Medline: 10761776]
42. Hariharan M, Polat K, Yaacob S. A new feature constituting approach to detection of vocal fold pathology. Int J Syst Sci 2014;45(8):1622-1634. [doi: 10.1080/00207721.2013.794905]

43. Lopes LW, Batista Simões L, Delfino da Silva J, da Silva Evangelista D, da Nóbrega E Ugulino AC, Oliveira Costa Silva P, et al. Accuracy of acoustic analysis measurements in the evaluation of patients with different laryngeal diagnoses. J Voice 2017 May;31(3):382.e15-382.e26. [doi: 10.1016/j.jvoice.2016.08.015] [Medline: 27742492]
44. Mohammed MA, Abdulkareem KH, Mostafa SA, Khanapi Abd Ghani M, Maashi MS, Garcia-Zapirain B, et al. Voice pathology detection and classification using convolutional neural network model. Appl Sci 2020 May 27;10(11):3723. [doi: 10.3390/app10113723]
45. Souissi N, Cherif A. Artificial neural networks and support vector machine for voice disorders identification. Int J Adv Comput Sci Appl 2016;7(5):339-344. [doi: 10.14569/ijacsa.2016.070546]
46. Wang X, Zhang J, Yan Y. Discrimination between pathological and normal voices using GMM-SVM approach. J Voice 2011 Jan;25(1):38-43. [doi: 10.1016/j.jvoice.2009.08.002] [Medline: 20137892]
47. Systematic reviews: CRD's guidance for undertaking reviews in healthcare. Centre for Reviews and Dissemination. 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2022-06-12]
48. Hariharan M, Polat K, Sindhu R, Yaacob S. A hybrid expert system approach for telemonitoring of vocal fold pathology. Appl Soft Comput 2013 Oct;13(10):4148-4161. [doi: 10.1016/j.asoc.2013.06.004]
49. Mehta DD, Hillman RE. Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. Curr Opin Otolaryngol Head Neck Surg 2008 Jun;16(3):211-215 [FREE Full text] [doi: 10.1097/MOO.0b013e3282fe96ce] [Medline: 18475073]
50. Mesallam TA, Farahat M, Malki KH, Alsulaiman M, Ali Z, Al-Nasheri A, et al. Development of the Arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. J Healthc Eng 2017;2017:8783751 [FREE Full text] [doi: 10.1155/2017/8783751] [Medline: 29201333]
51. Arias-Londoño JD, Godino-Llorente JI, Sáenz-Lechón N, Osma-Ruiz V, Castellanos-Domínguez G. An improved method for voice pathology detection by means of a HMM-based feature space transformation. Pattern Recognit 2010 Sep;43(9):3100-3112. [doi: 10.1016/j.patcog.2010.03.019]
52. Harar P, Galaz Z, Alonso-Hernandez JB, Mekyska J, Burget R, Smekal Z. Towards robust voice pathology detection. Neural Comput Applic 2018 Apr 04;32(20):15747-15757. [doi: 10.1007/s00521-018-3464-7]

Abbreviations

DTA: diagnostic test accuracy
K-NN: K-nearest neighbor
LS-SVM: least-squares support-vector machine
MEEI: Massachusetts Eye and Ear Infirmary
ML: machine learning
QDA: quadratic discriminant analysis
QUADAS: Quality Assessment of Diagnostic Accuracy Studies
SLT: speech and language therapist
UPM: Universidad Autónoma de Madrid

Edited by R Kukafka; submitted 04.04.22; peer-reviewed by A Ahmed, D Alhuwail; comments to author 22.04.22; revised version received 17.06.22; accepted 28.07.22; published 14.10.22.

Please cite as:
Al-Hussain G, Shuweihdi F, Alali H, Househ M, Abd-alrazaq A
The Effectiveness of Supervised Machine Learning in Screening and Diagnosing Voice Disorders: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(10):e38472
URL: https://www.jmir.org/2022/10/e38472
doi:10.2196/38472
PMID:36239999

Review

Credibility, Accuracy, and Comprehensiveness of Readily Available Internet-Based Information on Treatment and Management of Peripheral Artery Disease and Intermittent Claudication: Review

Shelley Alexander[1*], BSc, MSc; Chris Seenan[1*], BSc, PhD

Department of Physiotherapy and Paramedicine, School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Chris Seenan, BSc, PhD
Department of Physiotherapy and Paramedicine
School of Health and Life Sciences
Glasgow Caledonian University
Cowcaddens Road
Glasgow, G4 0BA
United Kingdom
Phone: 44 1413318151
Email: chris.seenan@gcu.ac.uk

Abstract

Background: Peripheral artery disease (PAD) affects millions of people worldwide, and a core component of management of the condition is self-management. The internet is an important source of health information for many people. However, the content of websites regarding treatment recommendations for PAD has not been fully evaluated.

Objective: This study aimed to assess the credibility, accuracy, and comprehensiveness of websites found via a common search engine, by comparing the content to current guidelines for treatment and management of PAD and intermittent claudication (IC).

Methods: A review of websites from hospitals, universities, governments, consumer organizations, and professional associations in the United States and the United Kingdom was conducted. Website recommendations for the treatment of PAD and IC were coded in accordance with the guidelines of the National Institute for Health and Care Excellence (NICE) and the American Heart Association (AHA). Primary outcomes were website credibility (4-item Journal of the American Medical Association benchmark), website accuracy (in terms of the percentage of accurate recommendations), and comprehensiveness of website recommendations (in terms of the percentage of guideline recommendations that were appropriately covered). Secondary outcomes were readability (Flesch–Kincaid grade level) and website quality (Health On the Net Foundation's code of conduct).

Results: After screening, 62 websites were included in this analysis. Only 45% (28/62) of websites met the credibility requirement by stating they were updated after the NICE guidelines were published. Declaration of authorship and funding and the presence of reference lists were less commonly reported. Regarding accuracy, 81% (556/685) of website recommendations were deemed accurate on following NICE's and the AHA's recommendations. Comprehensiveness was low, with an average of 40% (25/62) of guideline treatment recommendations being appropriately covered by websites. In most cases, readability scores revealed that the websites were too complex for web-based consumer health information.

Conclusions: Web-based information from reputable sources about the treatment and management of PAD and IC are generally accurate but have low comprehensiveness, credibility, and readability.

(J Med Internet Res 2022;24(10):e39555) doi:10.2196/39555

KEYWORDS

peripheral artery disease; intermittent claudication; health information; education; internet; eHealth; digital health

Introduction

The internet is increasingly being used by the general public as a source of health information [1]. People may use an internet search at various times along a health care journey: prior to seeking medical advice from their health care providers, to support self-management, and to make treatment decisions [1]. This is especially relevant to people with peripheral arterial

disease (PAD), where self-management and behavior change are key aspects of care [2]. Clinical guidelines that summarize the best available research evidence and expert consensus for the diagnosis and management of PAD have been developed by the American Heart Association (AHA) [3] and the National Institute of Health and Care Excellence (NICE) [4]. These guidelines recommend lifestyle modifications including cessation of smoking, a healthy diet, sustaining a healthy weight, and regular physical activity. Prescriptions of antiplatelets, statins, antihypertensives, and vasodilators are also recommended, but stenting and bypass surgery should only be considered if structured exercise and lifestyle modifications have been exhausted [3,4].

Most commonly, search engines, such as Google, are used as the method of searching for health information on the internet [5]. A 2014 report found that 60% of UK respondents had used the internet to search for health information in the previous 12 months, with younger people being more likely to search for information in this manner than older generations [1]. It is likely for these figures to have increased, especially considering the global COVID-19 pandemic. A recent survey found that health services had been completely or partially disrupted in many countries as a result of the pandemic, including services for hypertension, diabetes, cancer, and cardiovascular emergencies [6]. This indicates that potentially less health care provision for people with PAD may have been available since the onset of the pandemic. Along with reductions in access to in-person health care, the number of people searching the internet for health care–related information, such as that on PAD, could have increased. It is also likely for older generations to use the internet more than they did previously, and as this is the demographic more likely to experience PAD, there may be more people searching for such information on the internet than ever before.

Owing to the importance of self-management in long-term conditions and the expanding role of the internet in gathering health information, it is essential that web-based sources of information are accurate, credible, and comprehensive [7]. However, despite the large number of people seeking web-based health information, previous research has found the quality of these websites to be relatively poor [8-11]. To date, only one study has assessed the quality of information about PAD and intermittent claudication (IC) on websites and videos [12], and no previous research has compared web-based information to clinical guidelines. By assessing the quality of this information and any gaps or inaccuracies within it, recommendations can be made to improve the quality of information on the internet while optimizing the quality of life for those living with PAD. Through accurate self-management advice and support with behavior change, internet searching could empower individuals to assume a more active role in managing their condition [10].

This review aims to compare trustworthy websites to current clinical guidelines for the treatment and management of PAD and IC to assess their credibility, accuracy, comprehensiveness, and readability.

Methods

Study Design

This review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines where possible [13]. A review of “trustworthy” websites from governments, hospitals, universities, professional bodies, and health care organizations was conducted.

Eligibility Criteria

Websites were sought from the United States and the United Kingdom and had to be written in English to be included. Websites were deemed trustworthy if they were from the government, nonprofit organizations, hospitals, universities, professional societies, or consumer organizations. To be included, websites had to mention at least one recommendation for the treatment or management of PAD or IC. Web links from Google searches, which directed us to PDFs, were included if they met every other criterion mentioned above. Websites were excluded if they were not freely accessible, required sign-in details, or required payment to be accessed. Any Google Ads on the 2 pages screened were excluded. Web links to other parts of the same website were followed and included, but web links leading to external sources were excluded. Websites were excluded if they were scientific journal articles, blogs, videos, or the comparative guidelines themselves.

Search Strategy

A recent review has indicated that although people seeking health information on the web tend to use health websites, they often start with generic search engines, most commonly Google [5]. Google was used to search for freely accessible, noncommercial websites presenting information on PAD or IC during October 21-23, 2020, and updated during October 1-3, 2021. Search terms were decided after some trial searching on Google Trends. It was found that abbreviations “PAD” and “IC” were often searched for reasons other than those for “peripheral artery disease” and “intermittent claudication,” respectively; hence, the nonabbreviated versions were used as search terms. To target more trustworthy websites including government, hospital, university, and consumer organization websites, various words were added after the initial search terms [10]. A full breakdown of search terms is shown in Table 1. The search engine, Google, was used as it is considered the most common search engine with the best search validity [14]. For increased search specificity, each term was searched on both the UK (google.co.uk) and US (google.com) domains of Google. The first 2 pages of results were screened from each search in line with the eligibility criteria. The browsing data were cleared between each search. All web links deemed relevant by the first reviewer (SA) were collated to an Excel (Microsoft Inc) spreadsheet and then screened for eligibility by the second reviewer (CS), with all discrepancies resolved through discussion.

Table 1. Search terms used on Google; adapted from Ferreira et al [10].

Peripheral artery disease	Intermittent claudication
United Kingdom (google.co.uk)	
Peripheral Artery Disease gov uk	*Intermittent Claudication gov uk*
Peripheral Artery Disease org uk	*Intermittent Claudication org uk*
Peripheral Artery Disease hospital uk	*Intermittent Claudication hospital uk*
peripheral Artery Disease university uk	*Intermittent Claudication university uk*
Peripheral Artery Disease association society uk	*Intermittent Claudication association society uk*
Peripheral Artery Disease consumer reports uk	*Intermittent Claudication consumer reports uk*
United States (google.com)	
Peripheral Artery Disease gov usa	*Intermittent Claudication gov usa*
Peripheral Artery Disease org usa	*Intermittent Claudication org usa*
Peripheral Artery Disease hospital usa	*Intermittent Claudication hospital usa*
peripheral Artery Disease university usa	*Intermittent Claudication university usa*
Peripheral Artery Disease association society usa	*Intermittent Claudication association society usa*
Peripheral Artery Disease consumer usa	*Intermittent Claudication consumer usa*

Data Extraction

Both reviewers (SA and CS) extracted data into separate spreadsheets then met to discuss and cross-check the data. Recommendations for PAD and IC treatment or management from each website were coded in accordance with the 2012 NICE recommendations (last updated in 2018) and the 2016 AHA guidelines for PAD and IC management and treatment [3,4]. There were minimal recommendations mentioned in one guideline but not in the other, and there were no conflicting recommendations. Each website recommendation was coded against the guidelines as endorsed by at least one guideline or dismissed by at least one guideline [10]. Treatment and management recommendations from websites were each compared to the combined guideline recommendations and coded, as seen in Table 2.

Table 2. Code for which websites were compared and graded [10].

Coding criteria	Description
Appropriate endorsement	A website recommendation to use a treatment that was also endorsed by at least 1 guideline.
Appropriate dismissal	A website recommendation to avoid a treatment that was also dismissed by at least 1 guideline.
Inappropriate endorsement	A website recommendation to use a treatment that was dismissed by at least 1 guideline.
Inappropriate dismissal	A website recommendation to avoid a treatment that was endorsed by at least 1 guideline.
Endorsed	A website recommendation to use a treatment not mentioned in either guideline.
Dismissed	A website recommendation to avoid a treatment not mentioned in either guideline.
Unclear	A website recommendation that was too vague to be clearly matched to the guidelines or led to discrepancies between researchers.

Outcomes

Credibility

The credibility of each website was assessed using the Journal of the American Medical Association (JAMA) benchmark [10,15]. The JAMA benchmark evaluates websites on 4 items: (1) information currency, (2) authorship declaration, (3) presence of a reference list, and (4) disclosure of any conflicts of interest, sponsorship, or funding. Information was deemed current if it was dated after NICE guidelines were published (August 8, 2012) [16]. A declaration of authorship was included if single or multiple authors were mentioned, or authorship was tied to a group or entity [10]. Each of the 4 items was answered with "Yes," "No," or "Not reported."

Accuracy

The number of recommendations from websites that were accurate and clear were defined as those that were coded as appropriate endorsements, appropriate dismissals, or dismissed treatments not mentioned in either guideline. Recommendations were deemed inaccurate if they were coded as inappropriate endorsements, inappropriate dismissals, or endorsed treatments not mentioned in either guideline [10].

Comprehensiveness

The proportion of accurate guideline recommendations covered by a website was determined to measure their comprehensiveness. Website comprehensiveness was determined from the ratio of the sum of appropriate endorsements and

dismissals against the total number of recommendations in the comparative guidelines [10].

Readability

The Flesch–Kincaid grade level (FKGL) [17] is widely accepted as an appropriate instrument to evaluate the readability of general and health documents. The FKGL yields a reading level score calculated from the average length of words and sentences in a discourse. The score yielded by the FKGL is rated against US school levels, and it has been suggested that the FKGL should be between 6 and 8 for medical and health information aimed at the general public [18,19]. In this study, the websites were segregated into 3 groups based on the FKGL: <8, 8-10, and >10. Websites with an FKGL of <8 are deemed accessible to most people, those scoring 8-10 are accessible to some, and those scoring >10 are deemed inaccessible to the majority of the UK or US readers. The FKGL was calculated in this study using the inbuilt readability function in Word (version 2013; Microsoft Inc). A large section of text was copied from each website and pasted onto Word before running the readability statistic.

HONcode

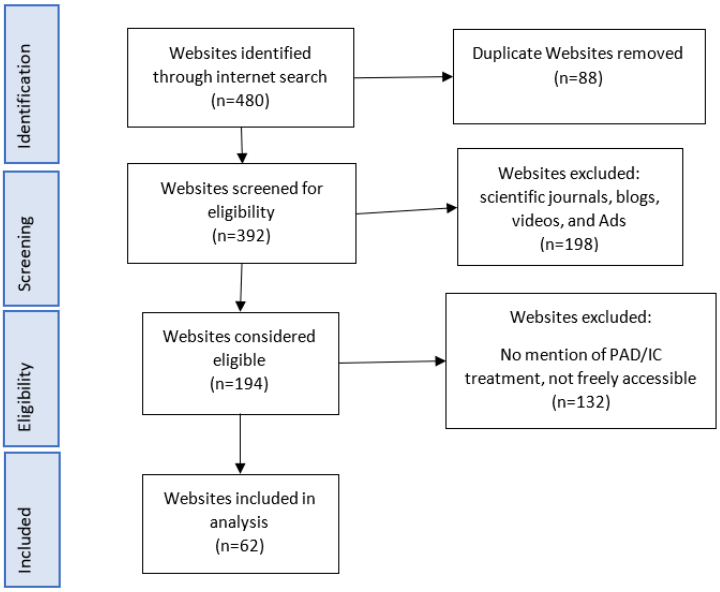
The Health On the Net (HON) Foundation's code of conduct (HONcode) is a well-known ethical and trustworthy code for evaluating the quality of medical and health information available on the internet. Websites that follow the HONcode principles can be approved by the HON foundation and be allowed to display the HONcode certificate symbol at the bottom of the page as a benchmark of quality. The HONcode has been used in similar previous studies as an indication of web-based health information quality [11,17,18]. In this review, the item was scored as "Yes" or "No" for presence of the HONcode logo on each website.

Results

Website Selection

Searches were conducted 24 individual times with the first 2 pages of results being assessed for eligibility. Searches collated 480 website results, and after duplicate results (n=88) were removed, the rest were screened for eligibility. Of these websites, 330 were deemed ineligible with reasons given in Figure 1. The remaining 62 websites were included in the analysis.

Figure 1. Study flow diagram with reasons for website exclusion. IC: intermittent claudication; PAD: peripheral artery disease.



Website Characteristics

From the UK- and US-specific Google sites searched, 48% (30/62) of eligible websites were found from the United Kingdom and 52% (32/62) of them from the United States. A large proportion of the analyzed websites (45%, n=28) were those of hospitals, followed by those of nongovernment organizations (23%, n=14), universities (13%, n=8), government organizations (10%, n=6), consumer organizations (5%, n=3) and, finally, professional associations or societies (5%, n=3). Information about the characteristics of the websites is provided in Table 3.

Table 3. Characteristics of websites and credibility data (N=62).

Descriptive and credibility variables	Websites, n (%)
Country	
United Kingdom	30 (48)
United States	32 (52)
Website type	
Government	6 (10)
Hospital	28 (45)
University	8 (13)
Consumer organization	3 (5)
Nongovernment organization	14 (22)
Professional association or society	3 (5)
Updated in accordance with NICE[a] guidelines (August 8, 2012)	
Yes	27 (44)
No	2 (3)
Not reported	33 (53)
Authorship declared	
Yes	17 (27)
No	45 (73)
Contains a reference list	
Yes	15 (24)
No	47 (76)
Disclosure of conflicts of interest or funding	
Yes	1 (2)
Not reported	61 (98)

[a]NICE: National Institute for Health and Care Excellence.

Credibility

The date of publication or last review was present on 32 websites with 5 (8%) of these dated before the NICE guidelines were published and the other 27 (44%) dated after and therefore deemed as being up to date. However, 30 (48%) websites did not report a date on their web page. Authorship was declared on only 18 (29%) websites, and 15 (24%) websites presented a reference list. Disclosure of any conflicts of interest, sponsorship, or funding was declared on 1 (2%) website (Table 3). More details on the assessment of website credibility can be found in Multimedia Appendices 1-3.

Accuracy

From the 62 websites analyzed, a total of 685 recommendations were recorded, with 556 (81.2%) being accurate, 10 (1.5%) inaccurate, and 114 (16.6%) unclear (Table 4). Most recommendations by websites were to use a treatment (n=589, 85.9%) rather than to avoid a treatment. The proportion of accurate recommendations was the highest from among UK searches (87.8%) in comparison to US searches (75.3%). Searches for IC yielded a higher proportion of accurate recommendations (86.9%) than searches for PAD (78.5%).

Further information on website recommendation accuracy is presented in Multimedia Appendices 1-3. The treatments most appropriately endorsed by websites were smoking cessation and cholesterol management (53/62, 86%), followed closely by angioplasty (n=52, 84%), physical activity (n=51, 82%), and blood pressure management (n=51, 82%). Least appropriately endorsed treatments included annual flu vaccine (0%) and exercise to maximal pain (n=6, 10%). Pentoxifylline was the most common treatment to be inappropriately endorsed by 5 (8%) websites, followed by anticoagulants (n=2, 3%). The most unclear recommendation was stenting, with 32 (52%) websites mentioning stenting but not in enough detail to match the comparative guidelines.

Importantly, none of the recommended treatments were inappropriately dismissed by any website. The most common website recommendation that was not mentioned in the guidelines was looking after mental well-being, which was mentioned in 6 (10%) websites. Website recommendations to avoid treatments not mentioned in the guidelines included the following: avoiding cold temperatures, not wearing compression stockings, and avoiding medication or herbal remedies that have been deemed ineffective or dangerous.

Table 4. Accuracy of website recommendations for the treatment of peripheral artery disease or intermittent claudication.

Search terms	Recommendations, n	Unclear recommendations, n	Accurate recommendations, n	Accurate endorsements, n	Accurate dismissals, n
Peripheral Artery Disease UK	208	31	178	173	2
Peripheral Artery Disease USA	262	49	191	187	0
Intermittent Claudication UK	113	12	104	99	2
Intermittent Claudication USA	102	22	83	75	4
Total	685	114	556	534	8

Comprehensiveness

Overall comprehensiveness of the included websites was low, covering 38% of recommended guidelines, on average, with approximately 8 out of 21 accurate recommendations (Table 5). The most comprehensive website had 13 recommendations that clearly and accurately matched the comparative guidelines, resulting in a comprehensiveness of 62%. Ranging from 2 to 13 accurate recommendations, the comprehensiveness of the websites found was extremely varied (10%-62%). No website mentioned all recommended treatments from the guidelines and most mentioned less than half. Full details on comprehensiveness are provided in Multimedia Appendices 1-3.

Table 5. Comprehensiveness of recommendations for peripheral artery disease or intermittent claudication treatment by websites when compared to the guidelines of the National Institute for Health and Care Excellence and the American Heart Association.

	Guideline recommendations, n	Guideline recommendations accurately covered by websites, mean (SD; %[a])
Recommendations to use a treatment	16	8.7 (3.1; 41.4)
Recommendations to avoid a treatment	5	0.2 (0.6; 0.9)
Total treatment recommendations	21	8.9 (3.7; 42.3)

[a]Percentage of total guideline recommendations.

HONcode and Readability

Only 5 of 62 (8%) websites were found to have the HONcode logo displayed on their web page as a marker of website quality. Of the 62 websites, 3 (5%) had an FKGL of <8 as recommended for health information aimed at the general public [18]. An additional 17 (27%) websites had an FKGL of 8-10, and 42 (68%) websites scored >10. The FKGL for most of the websites (68%) is deemed too high, which would make it difficult for most of the population to comprehend the presented information (Table 6). The FKGL scores ranged from 5.7 to 16.4, which covers a vast range of reading levels.

Table 6. Website quality and readability results.

Evaluation instrument	Websites, n (%)
Health On the Net Foundation's code of conduct	
Yes	5 (8.1)
No	57 (91.9)
Flesch–Kincaid grade level	
<8	3 (4.8)
8-10	17 (27.4)
>10	42 (67.7)

Discussion

Principal Findings

This is the first study to compare web-based information from trustworthy sources for people with PAD to current clinical guidelines to assess their credibility, accuracy, and comprehensiveness. Website recommendations for the treatment or management of PAD and IC were found to have low credibility when measured against the JAMA benchmark. Most recommendations provided were accurate; however, most websites lacked comprehensiveness and were not always clear in their recommendations. A high proportion of websites were too difficult for the average person to read and thus understand the recommendations they provided.

As this is the first study to assess web-based information regarding PAD and IC with respect to NICE and AHA guidelines, no direct comparisons to previous literature can be made. However, the 81.2% of accurate recommendations by websites found in this study is higher than that reported in a similar study on low back pain and pancreatic cancer, where

only 43.3% and 55% of website treatment recommendations, respectively, were accurate [9,10]. Both of these studies included more results from their searches—the first 50 or 100 results from each search—than this study, which only included the first 20 results [10]. Screening more results may yield websites that are less related to the search terms on the latter pages and yield less accurate recommendations as a result. Individuals rarely look past the first 2 pages of search engine results; hence, screening the first 20 results (2 pages) will have covered the sites that people with PAD are most likely to view.

Even though the recommendation accuracy was high, comprehensiveness was low with websites, averaging 8 out of 21 accurate recommendations (38%) from the guidelines. This indicates that generally, websites do not go into enough depth about the variety of treatment options for PAD and IC. This finding is similar to that of a previous study, where websites covered 6.73 of 17 recommendations (40%) for low back pain on average [10].

Smoking is one of the strongest risk factors for PAD [20], and cessation in people with IC has been shown to reduce mortality [21]. This is reflected in the web-based information as smoking cessation was accurately recommended by 86% of websites. Lifestyle modifications are the first line of treatment for PAD and can reduce cardiovascular ischemic events and improve function [3]. Therefore, it is surprising that the next most appropriately endorsed recommendation from websites was angioplasty (n=52/62, 84%). Surgical procedures are not a first line of treatment for most people with PAD, but among the websites reviewed in this study, they are more commonly recommended than, for example, exercise. A large proportion of the analyzed websites were those of US hospitals (24/62, 39%), and these sites may be advocating more for the surgical services they provide. A study of web-based information on pancreatic cancer yielded similar findings, indicating that website recommendations from US treatment centers were focused on treatment options offered at their facilities [9]. While information provided on these websites is mostly accurate, it is not comprehensive enough and could introduce surgical bias, thus undermining the potential success of other management strategies.

The general lack of self-management information found on websites in this study is reflective of the overall attitudes toward PAD and IC treatment in both the United Kingdom and the United States. A recent review of patient experiences of PAD [22] reported that patients often have very limited understanding of their condition. Being unaware of the systemic nature of PAD while also lacking information on self-management techniques from health care professionals leads patients to believe that surgical interventions alone will “cure” them [23,24]. People with PAD are often not involved in treatment-related decision-making and believe that doctors and surgeons know best, leading to unrealistic expectations from surgical interventions [25]. These individuals do not consider walking as a treatment—this is an illustration of the limited education about their condition that they are receiving from health care professionals [25]. Patients often feel the need to seek further information from friends, family, and the internet, making it even more important for web-based information to be accurate and comprehensive [26].

This perhaps also highlights a wider issue related to education on and the management of PAD and IC. Health care professionals, often in the context of limited resources, may refer their patients to web-based information, who in turn believe that the sources are comprehensive and accurate. This may serve as a substitute for or to supplement the education provided in the clinic. The findings of this review indicate that even websites normally considered reliable—for example, those of the AHA, British Heart Foundation, and Mayo clinic (see Multimedia Appendices 1-3)—have substantial limitations. Further engagement of specialized clinicians and educators in developing, reviewing, and signposting educational resources is required and may contribute to improved knowledge even among health care professionals [27].

After smoking cessation, exercise is arguably the next most important recommended self-management treatment for PAD and IC [28]. Therefore, it is promising that many websites (51/62, 82%) recommended this accurately. The literature suggests that supervised exercise programs (SEPs) are more beneficial to people with PAD and those with IC than general advice on home exercise [29,30]. However, SEPs have been much less frequently and appropriately endorsed by websites (21/62, 34%). Even though SEPs are endorsed by the NICE and AHA guidelines [3,16], lack of resources and funding often prevent their widespread use in practice [31]. Therefore, recommendations on websites alone are not enough to improve the management of PAD. There needs to be cohesion among guidelines, website recommendations, and the availability of health care resources to allow the provision of optimal care for patients with PAD.

Worldwide, the HON Foundation is recognized as an organization that assesses the quality of web-based health information directed at patients. In this study, only a small proportion of websites presenting PAD and IC treatment recommendations displayed the HONcode certificate logo (8%), which is lower than that reported for websites providing information regarding idiopathic pulmonary fibrosis (15%) and low back pain (41%) [8,11]. Importantly, HON Foundation–certified websites were not drastically better or more or less readable than noncertified websites. They tended to include more than the average number of accurate recommendations (8/21) but still also included many unclear recommendations. This suggests that the HONcode does not completely reflect the quality or accuracy of websites providing health information, which adds to the challenge of determining the accuracy of web-based health information for patients.

A significant concern regarding web-based health information is how accessible this information is to the average reader. Previous research has found most health websites do not have acceptable readability levels, including those designed for people with PAD [12]. Studies assessing web-based information on inflammatory bowel disease and pancreatic cancer found that only 4%-5% of websites were “readable,” as revealed by a FKGL of <8 [9,18]. Similarly, only 5% of websites in this review achieved this acceptable reading level. The average

FKGL of websites supplying PAD and IC health information was 11.2, which is much higher than the grade 6-8 level recommended for this type of information. Increased accuracy of websites is associated with increased reading level scores, and it seems to be difficult to produce accurate and easily understood information for all audiences [32]; however, it is important that all those who provide medical information via web-based resources are aware of the importance of providing both accurate and readable content.

Limitations

In this study, the only search engine used was Google as it is known to have the best search validity and is the most popular search engine [10,15]. Multiple search engines have been used in other studies to enhance the likelihood of finding all relevant websites. The literature on this is conflicting; however, studies have shown that only 1% of first-page results were the same when searched on both Google and Yahoo [18], with a high degree of overlap between results from different search engines [10]. Furthermore, the findings of this review may be limited from a global perspective owing to only seeking websites presented in English and only from US- and UK-specific website domains. However, as we were comparing website recommendations to the NICE and AHA guidelines, it was appropriate to use websites from corresponding countries. Using specific search terms to target “trustworthy” websites could be a limitation as the average person searching for this information would be unlikely to use these specific search terms. However, this meant that the recommendations by the websites are more likely to be trusted and followed by individuals. In this study, ranking of search results and website layout and design were not evaluated, but it is likely that this may also affect a person's ability to access accurate information.

Conclusions

Websites recommending treatments and management of PAD and IC are mostly accurate but have low credibility, low comprehensiveness, and are too complex for the average person to understand. With an increasing number of individuals seeking health information on the internet, it is imperative that websites be of high quality and do not act as barriers to patient education or introduce bias or unrealistic expectations for care. Rather, they should support self-management and behavior change and should reflect the advice and treatment options provided by health care professionals. Websites presenting information on PAD and IC should do so in accordance with evidence-based guidelines as much as possible, and health care professionals must ensure that they are providing clear and complete information to people with PAD and IC to avoid them from lacking an understanding of their condition. Future research should further assess available web-based information on PAD and IC, as well as overall patient and professional perceptions of the condition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Credibility of websites as measured by the JAMA benchmark.
[DOCX File , 19 KB - jmir_v24i10e39555_app1.docx]

Multimedia Appendix 2
Frequency (%) of websites endorsing or dismissing treatments mentioned in NICE or AHA guidelines (n=62).
[DOCX File , 17 KB - jmir_v24i10e39555_app2.docx]

Multimedia Appendix 3
Accuracy and Comprehensiveness of website recommendations.
[DOCX File , 32 KB - jmir_v24i10e39555_app3.docx]

References

1. European Commission, Directorate-General for the Information Society and Media. European citizens' digital health literacy: report. Brussels: European Commission; 2015.
2. Behaviour change: general approaches. National Institute for Health and Care Excellence. 2007 Oct 24. URL: https://www.nice.org.uk/guidance/ph6 [accessed 2022-09-30]
3. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Circulation 2017 Mar 21;135(12):e686-e725 [FREE Full text] [doi: 10.1161/CIR.0000000000000470] [Medline: 27840332]
4. Lower limb peripheral arterial disease: diagnosis and management. National Clinical Guideline Centre. 2012. URL: https://www.nice.org.uk/guidance/cg147/evidence/full-guideline-pdf-186865023 [accessed 2022-09-30]
5. 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]

6. Brunier A, Harris M. COVID-19 significantly impacts health services for noncommunicable diseases. World Health Organization. 2020 Jun 01. URL: https://www.who.int/news/item/01-06-2020-covid-19-significantly-impacts-health-services-for-noncommunicable-diseases [accessed 2022-09-30]
7. Hesse B, Nelson D, Kreps G, Croyle R, Arora N, Rimer B, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
8. Hendrick PA, Ahmed OH, Bankier SS, Chan TJ, Crawford SA, Ryder CR, et al. Acute low back pain information online: an evaluation of quality, content accuracy and readability of related websites. Man Ther 2012 Aug;17(4):318-324. [doi: 10.1016/j.math.2012.02.019] [Medline: 22464886]
9. De Groot L, Harris I, Regehr G, Tekian A, Ingledew P. Quality of online resources for pancreatic cancer patients. J Cancer Educ 2019 Apr;34(2):223-228. [doi: 10.1007/s13187-017-1290-8] [Medline: 29047015]
10. Ferreira G, Traeger AC, Machado G, O'Keeffe M, Maher CG. Credibility, accuracy, and comprehensiveness of internet-based information about low back pain: a systematic review. J Med Internet Res 2019 May 07;21(5):e13357 [FREE Full text] [doi: 10.2196/13357] [Medline: 31066689]
11. Fisher JH, O'Connor D, Flexman AM, Shapera S, Ryerson CJ. Accuracy and reliability of internet resources for information on idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2016 Jul 15;194(2):218-225. [doi: 10.1164/rccm.201512-2393OC] [Medline: 26849779]
12. Dar T, Chowdhury MM, Coughlin PA. An assessment of available information on the world wide web for patients with lower limb arterial disease. Eur J Vasc Endovasc Surg 2021 Apr;61(4):620-627 [FREE Full text] [doi: 10.1016/j.ejvs.2021.01.002] [Medline: 33589327]
13. 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. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
14. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012 May 16;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
15. Bernstam EV, Shelton DM, Walji M, Meric-Bernstam F. Instruments to assess the quality of health information on the world wide web: what can our patients actually use? Int J Med Inform 2005 Jan;74(1):13-19. [doi: 10.1016/j.ijmedinf.2004.10.001] [Medline: 15626632]
16. Peripheral arterial disease: diagnosis and management. London: National Institute for Health and Care Excellence (NICE); 2012.
17. Choudhry AJ, Baghdadi YMK, Wagie AE, Habermann EB, Heller SF, Jenkins DH, et al. Readability of discharge summaries: with what level of information are we dismissing our patients? Am J Surg 2016 Mar;211(3):631-636 [FREE Full text] [doi: 10.1016/j.amjsurg.2015.12.005] [Medline: 26794665]
18. van der Marel S, Duijvestein M, Hardwick JC, van den Brink GR, Veenendaal R, Hommes DW, et al. Quality of web-based information on inflammatory bowel diseases. Inflamm Bowel Dis 2009 Dec;15(12):1891-1896. [doi: 10.1002/ibd.20976] [Medline: 19462423]
19. Barnett T, Hoang H, Furlan A. An analysis of the readability characteristics of oral health information literature available to the public in Tasmania, Australia. BMC Oral Health 2016 Mar 17;16:35 [FREE Full text] [doi: 10.1186/s12903-016-0196-x] [Medline: 26984514]
20. Willigendael EM, Teijink JAW, Bartelink M, Kuiken BW, Boiten J, Moll FL, et al. Influence of smoking on incidence and prevalence of peripheral arterial disease. J Vasc Surg 2004 Dec;40(6):1158-1165. [doi: 10.1016/j.jvs.2004.08.049] [Medline: 15622370]
21. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. Circ Res 2015 Apr 24;116(9):1509-1526. [doi: 10.1161/CIRCRESAHA.116.303849] [Medline: 25908725]
22. Abaraogu UO, Ezenwankwo EF, Dall PM, Seenan CA. Living a burdensome and demanding life: A qualitative systematic review of the patients experiences of peripheral arterial disease. PLoS One 2018;13(11):e0207456 [FREE Full text] [doi: 10.1371/journal.pone.0207456] [Medline: 30440040]
23. Lokin JL, Hengeveld PJ, Conijn AP, Nieuwkerk PT, Koelemay MJ. Disease understanding in patients with intermittent claudication: a qualitative study. J Vasc Nurs 2015 Sep;33(3):112-118. [doi: 10.1016/j.jvn.2015.02.002] [Medline: 26298615]
24. Gibson JM, Kenrick M. Pain and powerlessness: the experience of living with peripheral vascular disease. J Adv Nurs 1998 Apr;27(4):737-745. [doi: 10.1046/j.1365-2648.1998.00599.x] [Medline: 9578203]
25. Galea Holmes MN, Weinman JA, Bearne LM. 'You can't walk with cramp!' A qualitative exploration of individuals' beliefs and experiences of walking as treatment for intermittent claudication. J Health Psychol 2017 Feb;22(2):255-265. [doi: 10.1177/1359105315600238] [Medline: 26349613]
26. Gorely T, Crank H, Humphreys L, Nawaz S, Tew GA. "Standing still in the street": experiences, knowledge and beliefs of patients with intermittent claudication--a qualitative study. J Vasc Nurs 2015 Mar;33(1):4-9. [doi: 10.1016/j.jvn.2014.12.001] [Medline: 25700732]

27. Bridgwood BM, Nickinson AT, Houghton JS, Pepper CJ, Sayers RD. Knowledge of peripheral artery disease: What do the public, healthcare practitioners, and trainees know? Vasc Med 2020 Jun;25(3):263-273. [doi: 10.1177/1358863X19893003] [Medline: 32000617]
28. McDermott MM. Exercise training for intermittent claudication. J Vasc Surg 2017 Nov;66(5):1612-1620 [FREE Full text] [doi: 10.1016/j.jvs.2017.05.111] [Medline: 28874320]
29. Olin JW, Sealove BA. Peripheral artery disease: current insight into the disease and its diagnosis and management. Mayo Clin Proc 2010 Jul;85(7):678-692 [FREE Full text] [doi: 10.4065/mcp.2010.0133] [Medline: 20592174]
30. Hamburg NM, Creager MA. Pathophysiology of intermittent claudication in peripheral artery disease. Circ J 2017 Mar 24;81(3):281-289 [FREE Full text] [doi: 10.1253/circj.CJ-16-1286] [Medline: 28123169]
31. Harwood AE, Pymer S, Ibeggazene S, Ingle L, Caldow E, Birkett ST. Provision of exercise services in patients with peripheral artery disease in the United Kingdom. Vascular 2022 Oct 04;30(5):874-881 [FREE Full text] [doi: 10.1177/17085381211035259] [Medline: 34348503]
32. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the accuracy and readability of online health information for patients with pancreatic cancer. JAMA Surg 2016 Sep 01;151(9):831-837. [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]

Abbreviations

AHA: American Heart Association
FKGL: Flesch–Kincaid grade level
HON: Health On the Net
HONcode: Health On the Net Foundation's code of conduct
IC: intermittent claudication
JAMA: Journal of the American Medical Association
NICE: National Institute for Health and Care Excellence
PAD: peripheral artery disease
SEP: supervised exercise program

Edited by R Kukafka; submitted 13.05.22; peer-reviewed by A Bashir, H Mehdizadeh, H Parisod; comments to author 28.06.22; revised version received 12.07.22; accepted 28.09.22; published 17.10.22.

Please cite as:
Alexander S, Seenan C
Credibility, Accuracy, and Comprehensiveness of Readily Available Internet-Based Information on Treatment and Management of Peripheral Artery Disease and Intermittent Claudication: Review
J Med Internet Res 2022;24(10):e39555
URL: https://www.jmir.org/2022/10/e39555
doi:10.2196/39555
PMID:36251363

Review

eHealth in Care Coordination for Older Adults Living at Home: Scoping Review

Hilde Marie Hunsbedt Fjellså[1], MSc; Anne Marie Lunde Husebø[1,2], MSc, PhD; Marianne Storm[1,3], MSc, PhD

[1]Department of Public Health, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

[2]Research Group of Nursing and Health Sciences, Stavanger University Hospital, Stavanger, Norway

[3]Faculty of Health Sciences and Social Care, Molde University College, Molde, Norway

Corresponding Author:
Hilde Marie Hunsbedt Fjellså, MSc
Department of Public Health
Faculty of Health Sciences
University of Stavanger
Universitetet i Stavanger
Postboks 8600
Stavanger, 4036
Norway
Phone: 47 51832480
Email: hildemarie.fjellsa@uis.no

Abstract

Background: The population of older adults is projected to increase, potentially resulting in more older adults living with chronic illnesses or multimorbidity. Living with chronic illnesses increases the need for coordinated health care services. Older adults want to manage their illnesses themselves, and many are positive about using eHealth for care coordination (CC). CC can help older adults navigate the health care system and improve information sharing.

Objective: This study aimed to map the research literature on eHealth used in CC for older adults living at home. This study assessed CC activities, outcomes, and factors influencing the use of eHealth in CC reported by older adults and health care professionals.

Methods: We used a scoping review methodology. We searched four databases—MEDLINE, CINAHL, Academic Scoping Premier, and Scopus—from 2009 to 2021 for research articles. We screened 630 records using the inclusion criteria (older adults aged >65 years, primary health care setting, description of an eHealth program or intervention or measure or experiences with the use of eHealth, and inclusion of CC or relevant activities as described in the Care Coordination Atlas). The analysis of the included articles consisted of both a descriptive and thematic analysis.

Results: A total of 16 studies were included in this scoping review. Of these 16 studies, 12 (75%) had a quantitative design, and the samples of the included studies varied in size. The categories of eHealth used for CC among older adults living at home were electronic health records and patient portals, telehealth monitoring solutions, and telephone only. The CC activity communication was evident in all studies (16/16, 100%). The results on patient- and system-level outcomes were mixed; however, most studies (7/16, 44%) reported improved mental and physical health and reduced rehospitalization and hospital admission rates. Observing changes in patients' health was a facilitator for health care professionals using eHealth in CC. When using eHealth in CC, available support to the patient, personal continuity, and a sense of security and safety were facilitators for older adults. Individual characteristics and lack of experience, confidence, and knowledge were barriers to older adults' use of eHealth. Health care professionals reported barriers such as increased workload and hampered communication.

Conclusions: We mapped the research literature on eHealth-enabled CC for older adults living at home. We did not map the gray literature as we aimed to map the research literature (peer-reviewed research articles published in academic journals). The study results showed that using eHealth to coordinate care for older adults who live at home is promising. To ensure the successful use of eHealth in CC, we recommend customized eHealth-enabled health care services for older adults, including individualized education and support.

(J Med Internet Res 2022;24(10):e39584) doi:10.2196/39584

KEYWORDS

eHealth; care coordination; older adults; primary health care; mobile phone

Introduction

Background

It is estimated that the population of older adults aged >65 years will double between 2010 and 2050, and over half of them are expected to live with multimorbidity [1-4]. Aging causes older adults to live with potentially both frailty and chronic illnesses, both affecting their health trajectory [5]. Furthermore, noncommunicable diseases (cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes) are a global health challenge [6]. The World Health Organization calls for better management of noncommunicable diseases and mental health conditions in primary health care, especially among older adults [7]. Living with chronic illness or multimorbidity often results in fragmented health care services and a lack of information sharing among members of the health care team and between health care professionals and patients [8-10]. Care coordination (CC) can reduce system fragmentation, help patients navigate the health care system, and improve information sharing [11]. In this scoping review, we understand CC according to the definition by McDonald et al [12]: "The deliberate organization of patient care activities between two or more participants (including the patient) involved in a patient's care to facilitate the appropriate delivery of health care services [...]." We use the term *health care professionals*, which includes nurses working in both specialist and primary care, physicians, or general and specialist practitioners [13].

Health ITs (HITs), electronic health records (EHRs), and patient portals are important tools for CC that enable health care professionals and patients to share, access, and manage information [14,15]. Other types of eHealth include health applications, telehealth, contact through telephone use, and other medical devices such as sensor technology [16]. McDonald et al [11] describe HIT as an enabler of coordination as it makes it possible to exchange and share information and communicate among health care professionals as well as with patients [16].

Previous research has shown that many older adults want to manage their illnesses themselves [17-20], and many are positive about the use of eHealth [18]. This aligns with the expectation of treating and caring for older adults with multimorbidity or chronic illnesses in their homes [21-24]. An explorative qualitative study of primary health care professionals and older adults living at home points out that electronic care plans can improve primary care by ensuring accessible information for patients, next of kin, or health care professionals [25]. Husebø and Storm [26] reported that the use of video communication in web-based home visits to older adults can facilitate continuous and coordinated care between the patient and health care professionals. Improved information flow among health care professionals in the primary care and specialist health care services can be associated with fewer emergency department (ED) visits and reduce the likelihood of outpatient visits among older adults [27]. Kooij et al [28] conducted a systematic review of HIT interventions to support shared care for patients with chronic illnesses and reported that EHRs resulted in fewer rehospitalizations and more visits to primary care physicians.

Peterson et al [29] conducted a systematic scoping review of 37 CC frameworks and identified a need to increase the use of theoretical frameworks when assessing care initiatives, especially in a primary care setting. Peterson et al [29] pointed out that the definition of CC by McDonald et al [11] is the most cited. The Care Coordination Atlas framework focuses on organizing and evaluating measures [29]. McDonald et al [11] organized CC measures into activities that enhance CC. These activities are directed at health care professionals and include facilitating information exchange and communication, facilitating transitions, assessing the patient's needs and goals, creating a proactive plan of care, monitoring, following up and responding to change, supporting self-management goals, linking to community resources, and aligning resources with patient and population needs. The framework can adapt to the developing CC field and is especially relevant for CC in the primary health care setting [11]. The framework also suggests validated measures for each of the activities [11,29]. Thus, the Care Coordination Atlas was used in this scoping review to identify and report CC activities when using eHealth in CC for older adults.

Objectives

A limited amount of research has focused on eHealth to support CC in older adults [30]. There is a need to gain more knowledge on how best to support older adults using eHealth [31] and particularly to examine eHealth in CC for older adults living at home [30]. This scoping review mapped the research literature (peer-reviewed research articles published in academic journals) to explore the use of eHealth in CC for older adults. The research questions that guided our review were as follows: (1) What categories of eHealth can be identified in the research literature and how do the CC activities relate to the eHealth categories? (2) What are the patient and health care use outcomes associated with the use of eHealth in CC? (3) What factors influencing the use of eHealth in CC are reported by older adults and health care professionals?

Methods

Scoping Review Methodology

We followed the Arksey and O'Malley scoping studies framework [32]: (1) identifying the research questions; (2) identifying relevant studies; (3) selecting the studies; (4) charting the data; and (5) collating, summarizing, and reporting results. In addition, the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist and explanation developed by Tricco et al [33] were used as a reporting tool.

Identifying Relevant Studies (Databases and Search Terms)

The search was conducted in the MEDLINE, CINAHL, Academic Search Premier, and Scopus databases and included research articles published between 2009 and 2021. The last search was conducted in December 2021 by HMHF in collaboration with a university librarian. Search terms related to CC (*coordinated care, integrated care, integrated health, care management, patient care management, case management, care transition, continuity of care, care planning, continuum of care,* and *shared care*), eHealth (*telecare, telehealth, telemedicine, remote consultation, assistive technology, electronic health record, information communication technology,* and *mhealth*), home care (*home care services, home nursing, community-dwelling, independent living, home based care, community health services, municipal health services, primary health care,* and *general practitioner*), and older patients (*elderly, aged, older person, elderly and chronic illness, elderly and multimorbidity, older adult,* and *frail elderly*) were used. In addition, Medical Subject Headings and thesaurus terms were used when possible (see Multimedia Appendix 1 for all search terms and an example of a search).

Selection of Studies

The selected studies were included based on a 2-step iterative process. First, we developed and tested a set of preliminary eligibility criteria, which we used to screen the titles. All authors met to discuss the preliminary eligibility criteria and did some final modifications (Textbox 1). Second, we tested the final eligibility criteria on 20 titles and abstracts and found them fitting. The final eligibility criteria (Textbox 1) were used to screen all titles and abstracts in collaboration with all authors. HMHF screened all articles (both titles and abstracts), and AMLH and MS screened 30 titles and abstracts each. All full-text articles were screened by HMHF. The 3 authors screened the same 11 full-text articles. The included articles were read by all authors.

Textbox 1. Final eligibility criteria.

Inclusion criteria

- Older adults aged >65 years
- Primary health care setting; older adults living in their own home
- Describing an eHealth program or measure or intervention or experiences with the use of eHealth
- Including care coordination or relevant activities as described in the Care Coordination Atlas
- Published after 2009
- Reported in English
- Peer-reviewed when possible to choose a limitation in the database

Exclusion criteria

- Older adults aged <65 years, next of kin, informal caregivers, and studies including different age groups when it was not possible to extract data on those aged >65 years
- Older adults living in nursing homes or who were in a hospital
- Studies with a primary focus on cost-effectiveness
- Books, book chapters, literature reviews, study protocols, conference and poster abstracts and papers, editorials, and discussion papers

The EPPI-Reviewer (version 4; EPPI-Centre) software [34] was used in the screening process. All 3 authors met and discussed the inclusion and exclusion of records according to the eligibility criteria. Disagreements regarding inclusion or exclusion were resolved through discussions between the authors. Agreement on inclusion was reached for all articles.

Charting the Data

Descriptive data were charted from each article according to the following: authors; country of origin; study population (age group and number of participants); and type of eHealth program, intervention, measure, or experience with eHealth. For the articles that described patient or health care use outcomes, we extracted and charted these results when applicable. Data relevant to CC activities and factors influencing the use of eHealth in CC were also extracted. Data charting was conducted by HMHF with input from AMLH and MS.

Collating, Summarizing, and Reporting the Results

We prepared a descriptive summary of the study characteristics (country of origin, methods used, overview of included participants, and year of publication). We were inspired by a thematic analysis to thematically organize and present the study results [35]. The first author conducted an inductive analysis to identify codes of eHealth tools or solutions described in the articles, which were classified into 3 eHealth categories. Arksey and O'Malley [32] suggest using a theoretical framework to summarize and describe variables. Hence, we conducted a deductive thematic analysis to identify CC activities in the eHealth categories. HMHF searched for and documented relevant CC activities in the included studies. To identify patient and health care use outcomes and factors influencing the use of eHealth in CC, we used an inductive thematic analysis. Outcomes were coded and categorized into patient-level and system-level outcomes. When analyzing factors influencing the use of eHealth, HMHF identified codes, which were categorized

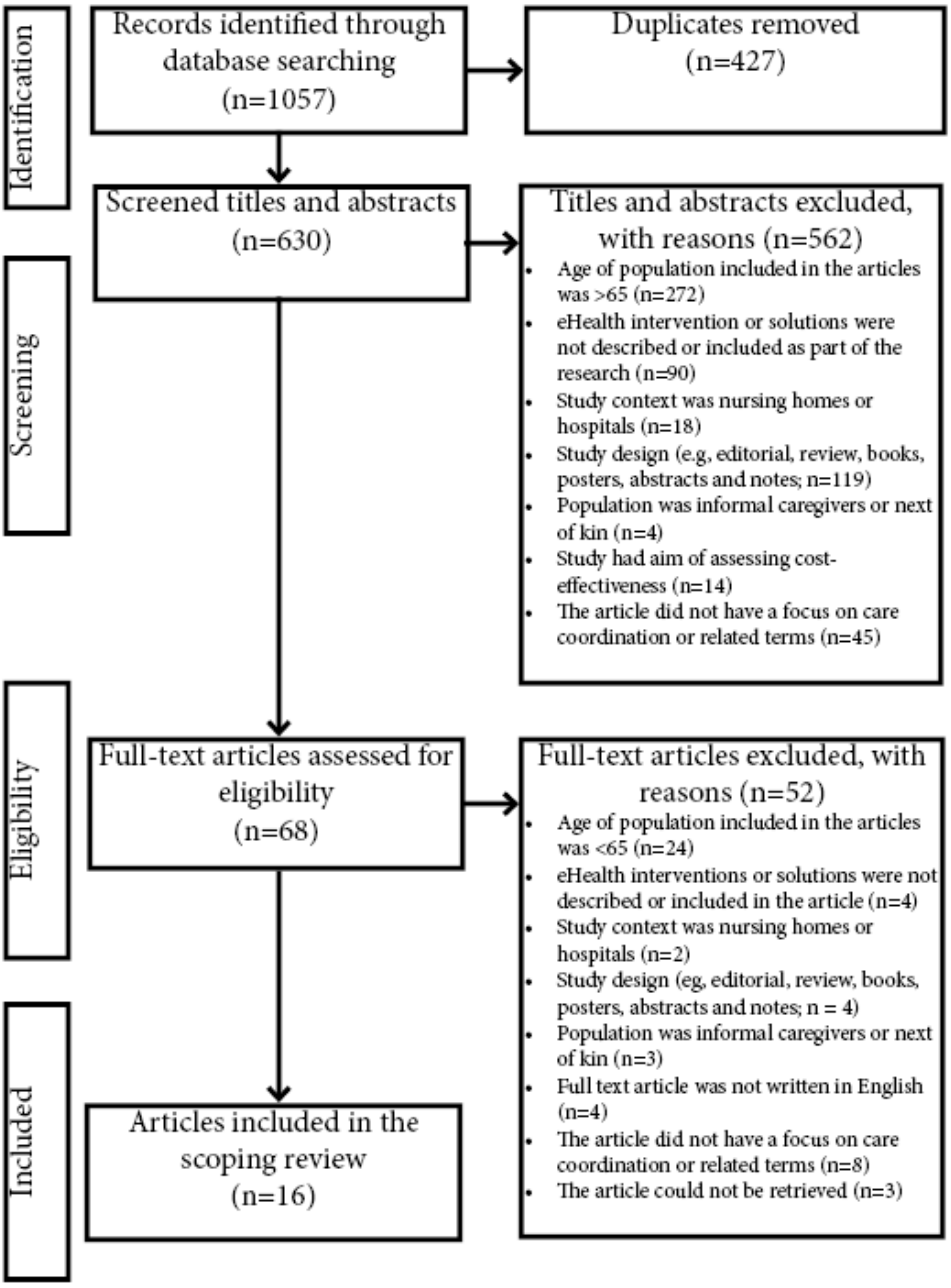
into facilitators of and barriers to the use of eHealth in CC. HMHF led the analysis process. The codes and identified categories were discussed with MS in 6 analysis meetings, and AMLH participated in 2 meetings.

Results

Overview

The study selection process is illustrated in Figure 1 [36]. A total of 1057 records were identified; after duplicates were removed, we screened the titles and abstracts of 630 (59.6%) records. A total of 89.2% (562/630) of titles and abstracts and 76% (52/68) of full-text articles were excluded, and the reasons are documented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram in Figure 1. The main reason for exclusion was that the study population did not meet the age criterion (>65 years). Another frequent reason for exclusion was that the study did not describe an eHealth intervention or experience with eHealth. A total of 68 articles were assessed in full text for eligibility, of which 16 (24%) were included in this scoping review.

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



Study Characteristics

Of the 16 included articles, 4 (25%) were from 2 research studies and had the same first authors: Makai et al [37,38] and Gellis et al [39,40]. The study sample sizes varied. The smallest sample size was reported in the study by Gokalp et al [41] (N=36). A research article presenting a randomized controlled trial (RCT) had the largest sample size (N=3661). See an overview of the study designs and participant characteristics in Table 1.

Table 1. Overview of study designs and participant characteristics (N=16).

Authors, country of origin	Type of study design	Participant characteristics	Perspectives represented in the study
Sheeran et al [42], United States	Quantitative pilot study	• 55 participants • 48 patients with depression • 7 health care professionals	• Patients • Health care professionals
Logue and Effken [43], United States	Quantitative pilot study	• 38 patients with chronic illnesses	• Patients
Gokalp et al [41], United Kingdom	Quantitative pilot study or technical review	• 36 patients; frail older adults with at least one chronic disease • Service team including health care professionals (number not specified)	• Patients • Health care professionals
Lewis et al [44], Ireland	Quantitative observational study	• 54 patients; frail older adults with comorbidities such as dementia, cardiovascular disease, hypertension, cerebral vascular disease, or COPD[a]	• Patients
De Jong et al [45], Netherlands	Quantitative observational study	• 96 patients with a dementia diagnosis	• Patients • Health care professionals
Makai et al [38], Netherlands	Quantitative controlled before-and-after study	• 682 patients; frail older adults • 290 patients in the intervention group • 392 patients in the control group	• Patients
Biese et al [46], United States	RCT[b]	• 120 patients who were discharged from the ED[c]; no requirements of chronic condition or diagnosis • 39 patients in the intervention group • 35 patients in the placebo group • 46 patients in the control group	• Patients • Health care professionals
Mavandadi et al [47], United States	RCT	• 1018 patients with depression or anxiety • 509 patients in the intervention group • 509 patients in the control group	• Patients
Gurwitz et al [48], United States	RCT	• 3661 patients who were discharged from the hospital (some had no diagnosis or chronic conditions, and others had diabetes, myocardial infarction, heart failure, COPD, cancer, stroke, cerebrovascular disease, or renal disease) • 1870 patients in the intervention group • 1791 patients in the control group	• Patients
Gellis et al [39], United States	RCT	• 115 patients with either heart failure or COPD and screened for depression • 57 patients in the treatment group • 58 patients in the control group	• Patients
Gellis et al [40], United States	RCT	• 115 patients with either heart failure or COPD and screened for depression • 57 patients in the treatment group • 58 patients in the control group	• Patients • Health care professionals
Cutrona et al [49], United States	Mixed methods; descriptive quantitative study and qualitative focus group study	• 799 patients who were discharged from the hospital (diagnosis not mentioned in the article) • Focus group with 5 physicians	• Patients • Health care professionals
Makai et al [37], Netherlands	Mixed methods; quantitative study and qualitative individual interviews	• 290 patients; frail older adults • 23 of these patients and their informal caregivers were included in semistructured individual interviews	• Patients • Health care professionals

Authors, country of origin	Type of study design	Participant characteristics	Perspectives represented in the study
Mateo-Abad et al [50], Spain	Mixed methods; quantitative and qualitative study	• 200 patients with two or more chronic conditions (at least one of them being COPD, chronic heart failure, or diabetes mellitus) • 101 patients in the intervention group • 99 patients in the control group • 9 qualitative interviews with patients, carers, clinicians, nurses, and managers	• Patients • Health care professionals
Dent and Tutt [51], United Kingdom	Longitudinal qualitative study	• 44 IT and health care professionals' experiences with implementation and integration of an IT-supported care pathway for frail older adults	• Health care professionals
Freilich et al [52], Sweden	Qualitative study	• 42 participants • 12 patients with multimorbidity (one chronic condition was either heart failure or diabetes mellitus) • 3 registered nurses • 20 physicians • 7 family caregivers	• Patients • Health care professionals

[a]COPD: chronic obstructive pulmonary disease.

[b]RCT: randomized controlled trial.

[c]ED: emergency department.

In total, 44% (7/16) of the studies were conducted in the United States [39,40,42,43,46-49]. A total of 12% (2/16) of the studies were conducted in the Netherlands [37,38,45], and 12% (2/16) were conducted in the United Kingdom [41,51]. In total, 6% (1/16) of the studies were conducted in each of the following three countries: Ireland [44], Spain [50], and Sweden [52]. Of the 16 articles, 7 (44%) were published in 2014 [35,37-39,41,45,50], and 4 (25%) were published in 2017 [44,49] and 2018 [41,45]. However, no studies were included from 2019 or 2021. A total of 12% (2/16) of the studies were published in 2020 [50,52].

Most studies (12/16, 75%) had a quantitative design [38-48] (Table 1). Among the 16 studies, there were 5 (31%) quantitative pilot or observational studies [41-45], 5 (31%) RCTs [39,40,46-48], and 1 (6%) before-and-after study [38]. A total of 19% (3/16) of the studies were mixed methods and combined qualitative and quantitative data [37,49,50], and 12% (2/16) had a qualitative design [51,52].

The included patients had a variety of chronic illnesses or multimorbidity, such as heart failure, chronic obstructive pulmonary disease, diabetes, cancer, dementia [39,40,44,45,48,50,52], frailty [37,38,41,44,51], or depression or anxiety [39,42,47]. A total of 56% (9/16) of the studies had participants who were older adults [37-40,43,44,46-48]. In total, 25% (4/16) of the studies were limited to older adults. However, this 25% (4/16) of studies also collected data on health care professionals' use of the eHealth intervention (how many times health care professionals opened alerts or accessed an electronic health portal) [37,40,45,46]. A total of 31% (5/16) of the studies included both older adults and health care professionals such as nurses, general practitioners, and health managers [41,42,49,50,52]. In total, 6% (1/16) of the studies were limited to health care professionals and focused on electronic integrated e-pathways for frail older adult patients [51].

Categories of eHealth and CC Activities

Overview

We identified three categories of eHealth in CC for older adults living at home in the included studies: (1) EHRs and patient portals, (2) telehealth monitoring solutions, and (3) telephone only. In the EHRs and patient portals category, electronic journals, personal health portals, and electronic personal health plans were used in CC [37,38,43,45,48-50]. In the telehealth monitoring solutions category, virtual ward or sensor technology was used [39-42,44,51,52]. In all these studies (7/7, 100%), sensor technology and telehealth monitoring were combined with electronic portals or home visits [39-42,44,51,52]. Telephones only were used in 12% (2/16) of the studies [46,47]. See Multimedia Appendix 2 [37-52] for an overview of the eHealth interventions or solutions used in the studies and the identified CC activities.

EHRs and Patient Portals

A total of 44% (7/16) of the studies [37,38,43,45,48-50] were classified in the EHRs and patient portals category. Logue and Effken [43] described barriers and facilitators when older adults used a personal health record to manage their health. De Jong et al [45] evaluated health care professionals' use of an electronic health portal (Congrendi). Makai et al [37,38] conducted an intervention on a health and welfare information portal (ZWIP) for frail older adults. Cutrona et al [49] explored the use of electronic messages sent upon hospital discharge of older adults and received by primary care physicians in the EHR. Mateo-Abad et al [50] conducted and evaluated an intervention including an electronic personal health folder with information, education, care plans, electronic messages between older adults (or their carers) and health care professionals, and monthly telephone and face-to-face meetings. Gurwitz et al [48] measured the effect of using EHRs and sent automatic alerts to

the primary care health care professionals when older adults were discharged from hospital to home.

In Table 2, CC activities according to the Care Coordination Atlas are described. Communication and information exchange is evident in 100% (7/7) of the studies. The study by Mateo-Abad et al [50] included several CC activities, for example, support for self-management goals where the patients were educated and guided on managing their chronic illness by using the health portal and over the telephone. Furthermore, the patients reported personal health data in the health portal [50]. In many studies (6/7, 86%), health care professionals received automatic alerts about information such as new medication, test results, or recommendations on treatment registered in the EHRs and patient portals [37,38,43,45,48,49]. Automatic alerts were related to the CC activities communication and information exchange, facilitation of transitions, monitoring, following up, and responding to change.

Table 2. Overview of Care Coordination Atlas activities in the electronic health records (EHRs) and patient portals category.

Care Coordination Atlas activities	EHRs and patient portals
Establish accountability or negotiate responsibility	• Patients were responsible for who they wanted to add to their electronic health portal [37,38]. • Patients had to give permission to begin a record and invite health care professionals to sign up. However, patients themselves did not use the eHealth portal [45].
Communicate	• Health care professionals and patients, or professionals and other health care professionals communicated with the help of electronic messages [45,49,50]. • Contact with the patient through telephone or in person [37,38,45,50] • Health care professionals received automatic digital alerts about relevant patient health information [37,38,43,45,48,50].
Facilitate transitions	• Primary care health care professionals received automated alerts when a patient was discharged from the hospital regarding discharge information, new drugs, medication warnings, and notification to schedule a follow-up appointment [48,49].
Assess needs and goals	• Patients could register care-related goals in the electronic care plan and initiate a change in the plan when a goal was reached [37,38].
Create a proactive plan of care	• By using and accessing an electronic personal health plan, patients were more involved and responsible for their health [50].
Monitor, follow up, and respond to change	• Health care professionals followed up on clinical information that was registered in the health portal [50].
Support self-management goals	• Patients received education and guidance on managing their chronic illness over the telephone or in the health portal [50].
Link to community resources	• Not evident
Align resources with patient and population needs	• Health care professionals in primary care experienced an increased workload with the new eHealth model [50].

Telehealth Monitoring Solutions

The category of telehealth monitoring solutions, such as sensor technology and virtual wards, was evident in 44% (7/16) of the studies [39-42,44,51,52]. Gokalp et al [41] focused on piloting a telemonitoring system for older adults, including various sensors such as pulse sensors, bed sensors, glucose meters, and blood pressure (BP) meters. Sheeran et al [42] tested the feasibility, acceptability, and clinical outcomes of a telemonitoring technology for older adults with depression. Lewis et al [44] monitored and tested a community ward integrating specialist and primary health care. Gellis et al [39,40] evaluated and examined the impact of a telehealth monitoring intervention, including a tabletop monitor at the homes of older adults where they could register weight, BP, pulse, and other vital signs. In addition, a health care professional conducted depression treatment sessions over the telephone for older adults with comorbid depression [40]. Dent and Tutt [51] reported health care professionals' experiences with an e-care pathway, including a virtual ward and telemonitoring of a patient in their home. Freilich et al [52] explored the perspectives of health care professionals, patients, and caregivers on the use of a telemedicine program, including telehealth monitoring (BP, weight, and blood sugar) and the use of a tablet to conduct video meetings with a nurse.

CC activities such as communicating and exchanging information, facilitating transitions, monitoring, following up, and responding to change were evident in all studies (7/7, 100%), as described in Table 3. In the studies by Lewis et al [44] and Dent and Trutt [51], the CC activity to facilitate transition was apparent as information on patient transfer was electronically sent from specialist to primary health care. Telephone and EHRs and patient portals were combined with telehealth monitoring, where health care professionals conducted education or counseling sessions over the telephone, video, or in EHRs and patient portals [39,40,42,52].

Table 3. Overview of Care Coordination Atlas activities in the telehealth monitoring solutions category.

Care Coordination Atlas activities	Telehealth monitoring solutions
Establish accountability or negotiate responsibility	• A telehealth nurse was assigned to be a care manager for the patient and contacted other health care professionals when necessary [42]. • A senior nurse was appointed as the clinical care manager and was responsible for patient care [44]. • Some patients reported not knowing if primary health care professionals or hospital specialists communicated with each other [52].
Communicate	• Education or counseling sessions were conducted over the telephone [39,40,42,52]. • The studies used a variation of home visits, video meetings, or telephone calls to patients or health care professionals [41,42,44,51,52]. • Patients' health data were registered in a portal and reviewed by a nurse [41,52].
Facilitate transitions	• Information about the patient was sent to primary health care when the patient was transferred between specialist and primary health care or needed a change in treatment [44,51]. • Telehealth nurses contacted and referred patients to primary care health care professionals when they observed changes in patients' health data [41]. • Different health care professionals were located together, and a care manager followed up with the patient across specialist and primary care [44]. • If a patient was discharged from hospital to home, a community nurse received an alert in an electronic portal and would ensure early discharge of the patient [51].
Assess needs and goals	• A telehealth nurse provided goal setting over the telephone with patients [42].
Create a proactive plan of care	• Not evident
Monitor, follow up, and respond to change	• Health care professionals monitored and assessed patient health data that were registered in an eHealth portal [39-42,44,51,52]. • Some patients felt secure knowing that a nurse kept track of their health parameters and would contact them if changes were observed [52].
Support self-management goals	• Patients received education or counseling sessions over the telephone or via an eHealth portal [39,40,42,52]. • Patients could ask questions or discuss a problem with a telehealth nurse when needed [39,40].
Link to community resources	• Not evident
Align resources with patient and population needs	• A virtual ward model with telehealth monitoring was set up with existing resources [44].

Telephone Only

A total of 12% (2/16) of the studies belonged to the third category, telephone only [46,47], and the telephone was used in combination with EHRs, patient portals, and telehealth monitoring solutions [39-42,44,50], as reported in Tables 2 and 3. As shown in Textbox 2, support for self-management goals was evident in the study by Mavandadi et al [47], where a nurse conducted symptom monitoring, education, and problem-focused therapy for older adult patients with depression and anxiety over the telephone. The study was an RCT; however, >20% of the included patients did not complete the intervention because of reduced cognitive function [47]. The RCT study by Biese et al [46] evaluated a telephone intervention in which older adults were telephoned within 5 days of an ED visit regarding making an appointment with physicians or medication changes. In this study, approximately 10% of the patients could not be reached by telephone after 3 attempts [46].

Textbox 2. Overview of Care Coordination Atlas activities in the telephone only category.

Care Coordination Atlas activities identified

- Establish accountability or negotiate responsibility: not evident
- Communicate: health care professionals contacted patients over the telephone [46,47]
- Facilitate transitions: a nurse telephoned patients 3 days after discharge from hospitals and helped patients who needed it navigate the health care system by reviewing discharge instructions and making appointments with or referrals to physicians [46]
- Assess needs and goals: not evident
- Create a proactive plan of care: not evident
- Monitor, follow up, and respond to change: health professionals monitored response to treatment and facilitated treatment over the telephone with patients [47]
- Support self-management goals: a study nurse conducted symptom monitoring, education, and problem-focused therapy with patients over the telephone [47]
- Link to community resources: not evident
- Align resources with patient and population needs: not evident

Patient-Level and System-Level Outcomes

Overall, 56% (9/16) of the studies measured the effect on patient- or system-level outcomes when implementing, piloting, or testing an eHealth solution [38-40,42,44,46-48,50]. See Table 4 for a detailed description of the interventions and outcomes. The patient-level outcomes were related to physical or mental health and social or problem-solving skills [38-40,42,47,50]. The system-level outcomes were related to health care use, such as hospitalizations, readmissions, follow-up visits with primary care health care professionals, or ED admission rates [39,44,46,48,50]. The patient-level outcomes were measured using standardized scales and survey questionnaires or recording vital signs throughout the intervention [38-40,42,47,50]. System-level outcomes were measured with objective scores, such as how often or if the patient went to the general practitioner or differences in hospitalization or ED visit rates between the intervention and control groups [39,44,46,48,50].

The patient- and system-level outcomes were mixed. Of the 9 studies, 7 (78%) showed improved physical or mental health [39,40,42,47,50], improved social and problem-solving skills [39,40], lower hospitalization rates, lower ED visits [39,44,46], and increased follow-up rates with primary care health care professionals [46,50]. In total, 11% (1/9) of the studies demonstrated no differences in physical or mental health between the intervention and control groups [38]. Makai et al [38] included 682 older adults in their study with a control and intervention group. The study by Gurwitz et al [48] did not demonstrate an increase in follow-up visits with primary care health care professionals or a reduction in rehospitalization rates. This study included >3661 patients but did not explore ways of communicating directly with patients or health care professionals other than sending automatic alerts to health care professionals [48].

Table 4. Overview of interventions and identified patient-level and system-level outcomes.

Intervention description	Patient-level outcomes	System-level outcomes
Mateo-Abad et al [50] conducted and evaluated the effect of an electronic personal health folder, which included accessing information, electronic messages, web-based education, monthly telephone calls, and face-to-face sessions with nurses. The intervention group used the electronic personal health folder and received usual care. The control group only received usual care. Outcomes related to clinical effect and the use of services were measured at two points throughout the intervention period (9 and 12 months). EHR[a] and administrative databases were used to extract available information.	Health data levels (BMI, blood pressure, blood glucose, and oxygen saturation) were significantly reduced in the intervention group compared with the control group [50].	There were lower hospitalization rate and increased appointments with general practitioners and nurses in the intervention group compared with the control group [50].
Makai et al [38] conducted a controlled before-and-after study of the health and welfare portal ZWIP. ZWIP contains a secure electronic messaging system and an EHR where the patient can invite health care professionals and their caregiver to join. Data were collected using a questionnaire with patients and their families at baseline and after 12 months. The control group received usual care.	The researchers observed no differences in physical or mental health between the intervention and control groups [38].	N/A[b]
The study by Gurwitz et al [48] assessed the effect of an EHR intervention. In the intervention group, automatic alerts were sent to the primary care health care professionals when older adults were discharged from the hospital. Data were collected on whether discharged individuals had an office visit with a primary care physician in the 7-, 14-, and 30-day periods after hospital discharge. The primary care health care professionals did not receive automatic alerts or information when older adults in the control group were discharged.	N/A	The study did not demonstrate an increase in follow-up visits with primary care health care professionals or a reduction in rehospitalization [48].
The quantitative pilot study by Sheeran et al [42] consisted of testing the feasibility, acceptability, and preliminary clinical outcomes of a telemonitoring technology to provide depression care. Data from older adults were collected at baseline and at the discharge of the intervention, which lasted a minimum of 3 weeks. Older adults had telephone contact or home visits by a telehealth nurse.	19 older adults had severe depression after the intervention, and 16 of them reported a mild depression score after the intervention [42].	N/A
Gellis et al [39] conducted an RCT[c] that tested the intervention, including the Honeywell Health Monitoring system. Weight, blood pressure, pulse, oxygen saturation, and temperature were monitored daily. A telehealth nurse was available for the older adults daily and monitored the data. Information from the older adults was collected using study questionnaires at baseline and at approximately 3 months. The control group received usual care.	Patients in the intervention group showed a greater increase in general health and social functioning than patients in the control group after 3 months [39].	The control group in their RCT study had significantly more visits to the ED[d] than the intervention group after 3 months [39].
Gellis et al [40] conducted an RCT that tested the intervention, including the Honeywell Health Monitoring system. In addition, the intervention group received chronic illness and depression care management and problem-solving treatment. A telehealth nurse monitored data and completed problem-solving treatment over the telephone with the older adults. A satisfaction survey, depression rating scale, and other information were collected at baseline and 3 and 6 months.	Results showed that the intervention group had greater problem-solving abilities, and their depression symptom scores improved significantly compared with those of the control group at the 3-month survey [40].	N/A
Lewis et al [44] conducted a quantitative observational study of a virtual ward using telehealth monitoring solutions. The virtual ward monitored older adults with home visits and telephone consultations. The risk of hospital admission was measured upon admission to the virtual ward. The number of unplanned admissions and ED presentations was measured before starting the intervention and upon discharge from the virtual ward.	N/A	The study demonstrated a reduction in ED visits and unplanned hospital admissions [44].
Mavandadi et al [47] conducted an RCT where the older adults in the intervention group received telephone-delivered symptom monitoring and were provided with educational and problem-focused therapy. The intervention group received maintenance calls at the 4-, 5-, and 6-month follow-ups. Both the control and the intervention group received 4 brief follow-up assessments over the telephone.	The older adults in the intervention group reported greater improvement in overall mental health functioning and reduced anxiety and depressive symptoms compared with those in the control group [47].	N/A

Intervention description	Patient-level outcomes	System-level outcomes
Biese et al [46] conducted an RCT that evaluated a telephone call intervention conducted by a trained nurse 1 to 3 days after ED discharge. The nurse followed a script and helped patients review discharge instructions and arranged appointments with physicians when needed. The placebo group received a satisfaction survey call 1 to 3 days after ED discharge, and the control group received no call. Telephone interviews were conducted with all groups 5 to 6 days and 30 to 35 days after ED discharge.	N/A	The older adults in the intervention group were more likely to see a physician within 5 days compared with the control and placebo groups [46]. The study further showed a reduction in the number of registered admissions in an ED; however, this was not significant compared with the control and placebo groups [46].

[a]EHR: electronic health record.

[b]N/A: not applicable.

[c]RCT: randomized controlled trial.

[d]ED: emergency department.

Facilitators of and Barriers to the Use of eHealth in CC

In the analysis of the articles, we identified two factors—facilitators and barriers—describing the use of eHealth in CC. A total of 8 facilitators and barriers were identified (see the overview in Textbox 3). Some of these barriers and facilitators were reported from patient satisfaction surveys [39,40,42]; descriptions of how health care professionals or older adults used the eHealth solution [46,47,50,51]; or qualitative data on experiences, evaluation, and use [37,41,49,50,52].

Textbox 3. Overview of facilitators of and barriers to the use of eHealth in care coordination.

Facilitators

- Available support to the patient
- Relation continuity between the older adult and health care professional
- A sense of security and safety
- New and valuable way to observe changes in patients' health

Barriers

- Individual characteristics
- Lack of experience, knowledge, or confidence regarding how to use eHealth
- Increased workload
- Hampered communication because of limited access to the electronic health records or patient portals

Available support to the patient was an important facilitator for older adults' use and management of eHealth technology [37,46,52]. Biese et al [46] reported that some older adults needed assistance to book appointments with health care professionals over the telephone. Makai et al [37] reported that some older adults had problems logging in to the electronic health portal, pointing out the importance of having available support to the patient. Freilich et al [52] claimed that some patients needed health care professionals to be in control and monitor their symptoms. Other patients did more of the monitoring and disease management themselves, making it necessary for health care professionals to tailor their support to the patients [52].

Relational continuity between the older adult and health care professional was important to facilitate the older adult's use of eHealth. In total, 12% (2/16) of the studies [50,52] reported that a close relationship between health care professionals and older adults supported the development and follow-up of electronic care plans. Another aspect highlighting the importance of relational continuity was that some older adults feared that eHealth would replace face-to-face contact, potentially negatively affecting eHealth use [35,52].

The use of eHealth in CC among older adults was also facilitated when they felt a sense of security and safety. A total of 19% (3/16) of the studies [37,38,50] reported that older adults felt reassured knowing that health care professionals were keeping track of their health data. In addition, being in charge of their health symptoms and communicating directly with health care professionals gave older adults a sense of safety and security [37,38,50].

Health care professionals' use of eHealth to coordinate care was facilitated when it was experienced as a new and valuable way of observing changes in patients' health [40,41,50]. In the study by Gellis et al [40], the nurses were attentive to changes in patient health by reviewing the health portal daily. Mateo-Abad et al [50] reported that the nurses who monitored the patient health data had greater familiarity with the older adults' chronic illnesses.

Individual characteristics such as being an older adult and having health problems such as hearing impairment and memory loss were barriers to the use of eHealth [38,43,47,50,52]. Adults aged ≥80 years did not use technology as often as younger older adults [38,43,47,51,52]. Mavandadi et al [47] reported that 23.6% of the 1018 included patients did not pick up the phone despite several attempts and having been given information about the study beforehand. According to Makai et al [38], most of the included patients rarely used the eHealth solution for coordination despite efforts from the researchers to implement and train them in its use.

The use of eHealth by older adults was also limited by a lack of experience, confidence, and knowledge of how to use it [41,43,52]. The patient group aged >80 years can be perceived as heterogenic and less confident in using technology than younger older adult patients [43]. Furthermore, Logue and Effken [43] found that more men than women expressed confidence in their ability to use technology.

Increased workloads limited health care professionals' use of eHealth. Some health care professionals experienced a heavier workload when implementing a new eHealth tool such as a patient portal [50]. In the studies by Biese et al [46] and Dent and Tutt [51], a dedicated nurse was in charge of facilitating transitions from hospital to home. The nurse arranged appointments and referrals and reviewed discharge instructions with the care team and the patient. Cutrona et al [49] found that primary care physicians perceived alerts in the EHR inbox as burdensome and, if the physicians had too many alerts in their inbox, the alert was less likely to be opened within 24 hours.

eHealth used for CC communication was hampered by limited access to the EHRs and patient portals for the health care professionals. De Jong et al [45] reported that less than half of the included patients had general practitioners linked to the EHR. The EHR was an additional system to what the general practitioners already used. A similar finding was reported in the studies by Makai et al [37,38], where patients could register care-related goals in a web-based care plan. In this study, not all health care professionals signed up to the portal or answered messages from the patients. In the study by Freilich et al [52], patient and family caregivers entered personal health data into the telehealth monitoring solution. However, this information was sent only to the primary health center. The primary health center belonged to the health region, but the home care nurses who visited the patients were employed by the municipality and did not have access to this information [52].

Discussion

Principal Findings

This review mapped the research literature on eHealth in CC for older adults living at home. We included 16 articles in the scoping review and identified three categories of eHealth: EHRs and patient portals, telehealth monitoring solutions, and telephone only. Communication was the CC activity reported in all the articles (16/16, 100%). Patient- and system-level outcomes were mixed. Most studies (7/16, 44%) showed that improved mental and physical health, reduced rehospitalization and hospital admissions, available support to the patient, relational continuity with health care professionals, and a sense of security were facilitators of older adults' use of eHealth in CC. Having new and useful tools for observing a change in patients' health facilitated health care professionals' use of eHealth in CC. Individual characteristics and lack of experience, confidence, and knowledge were barriers to older adults' use of eHealth in CC. Barriers reported by health care professionals were increased workload and hampered communication because of limited access to the EHRs and patient portals.

Comparison With Prior Work

We identified 3 eHealth categories when coordinating care for older adults. Despite the fast development of eHealth and technology, our results indicate that the telephone should still be considered necessary for older adults. A total of 12% (2/16) of the articles were classified under the category of telephone only [46,47]. However, the telephone was used in combination with EHRs, patient portals, and telehealth monitoring in 38% (6/16) of the studies [39-42,44,50]. Hawley et al [53] reported that, for older adults who were uninterested in and incapable of using eHealth, the telephone was important in the conduct of digital home visits. This is also supported by the study by Chu et al [54], where almost one-fifth of the older adults in the study did not have access to an electronic device, leaving the telephone as the only option to conduct virtual visits. EHRs and patient portals are commonly used in CC and integrated care programs, which is supported by other studies [11,55]. Melchiorre et al [55] categorize monitoring as an eHealth solution in integrated care programs, which supports the identification of the category of telehealth monitoring solutions.

Our results showed that communication was the dominant CC activity in all 3 eHealth types, a finding supported by other studies using the Care Coordination Atlas as a framework [56-58]. Both health care professionals and older adults communicated through electronic messages, over the telephone, via video, or in person. McDonald et al [11] highlighted that EHRs ensure information transfer between health care professionals. Our results document that facilitating transitions, supporting self-management goals, monitoring, following up, and responding to change are common CC activities in the 3 eHealth categories. Similarly, Chakurian and Popejoy [58] found the same CC activities when they used the CC framework to evaluate transitional care models. However, our review did not identify the CC activity community resources in the included eHealth types. This contrasts with the study by Samal et al [59], which reported the CC activity community resources as helpful regarding the automatic reference of patients to community programs when discharged from the ED or hospital. The Chronic Care Model also highlights that better access to community resources is important for individuals with chronic conditions [60].

A CC activity that appeared often was establishing accountability or negotiating responsibility [37,38,42,44,45,52]. Findings from our review show that older adults consented to who could access or start an electronic record in 19% (3/16) of the studies [37,38,45]. According to Tith et al [61], health care services still have challenges with patient consent. In Europe

and the United States, giving consent and overseeing who can access personal health information are included in policy standards such as the Health Insurance Portability and Accountability Act [53] and the General Data Protection Regulation [62]. However, Samal et al [59] stated that the activity of establishing accountability or negotiating responsibility has a low future potential to be used in HIT as it cannot be automated. However, the results of this scoping review show that, when using eHealth to coordinate care, it is essential to ensure that patients know what information is shared about them and with whom.

In total, 56% (9/16) of the studies measured the effect of the eHealth interventions. Some of the studies focusing on patient outcomes (5/6, 83%) showed greater social functioning and improved mental and physical health [39,40,42,47,50], indicating a greater quality of life [11]. McDonald et al [11] described that the end point of CC measures is, among other things, improved quality of life and reduced hospital readmissions and emergency room visits, which is in line with the patient and health care use outcomes of this scoping review.

We identified that older adults' characteristics, such as being very old and having health problems and memory loss, were barriers to eHealth use. Anderson and Perrin [63] reported that, even though more older adults than ever use smartphones in the United States, seniors aged 65 to 69 years are more likely to go on the web than those aged ≥80 years. Our results indicate that these older adults did not use technology as often as younger older adults, which can be described as the *digital divide* [63-65]. In addition to being older and less educated, impaired cognitive and numeracy ability, limited internet experience, and physical and visual impairment can limit the use of eHealth [66,67].

The digital divide can be explained in relation to eHealth literacy [68], where Rios et al [69] emphasized that training older adults in the use of technology can increase eHealth literacy. A recent mixed methods study by Fox and Connolly [70] found that it is important to educate older adults about mobile apps, wearable devices, and EHRs. Our results showed that available support to the patient and relational continuity could facilitate older adults' use of eHealth. Kim and Lee [71] reported limited information about training and support for patients when using electronic devices, which can hamper eHealth use [69]. Vroman et al [72] and Hawley et al [53] emphasized that training and education need to be personalized to the older adult's needs and skills. Sufficient technological support is also important to increase health literacy and narrow the digital divide when using eHealth [70,73].

A barrier that health care professionals reported was increased workloads and having new work tasks assigned related to the use of eHealth in CC. Gill et al [74] reported that health care professionals made significant efforts to gather patient information when using HIT to facilitate CC. According to Greenhalgh et al [75], new technologies can disrupt work processes, and some health care organizations cannot adapt to new ways of working.

The lack of interoperability across health systems hampers information exchange and communication when using eHealth in CC [9,11,74,76]. Hsiao et al [77] reported that office-based physicians who used HIT did not always receive the necessary patient information to coordinate care, especially from health care professionals outside their practice or hospital. Moreover, Liaw et al [76] addressed that the lack of a universal secure messaging system causes fragmented information sharing among health care professionals [9,59,77].

Future Directions

Our results showed that CC activities, including identifying community resources, establishing accountability, and negotiating responsibility, have a future potential for inclusion in eHealth research and practices. We recommend that community resources such as volunteer work, food delivery services, and support groups [11] be considered in both future research and practice. Furthermore, the digitalization of consent and responsibility can be enabled using a digital e-consent solution where patients can create, update, or withdraw their consent [61].

To narrow the digital divide and take into account the variety of older adults' individual characteristics, future researchers, practices, and policy makers need to consider using the telephone or in-person visits as a supplement or backup when conducting virtual visits or telehealth monitoring in CC. To meet the individual needs of older adults, customized support to the patient and education can be helpful to the successful use of eHealth in CC. Knowledge, confidence, and support are needed to ensure patient involvement when using eHealth to coordinate care. Therefore, future practice should have health care professionals with dedicated responsibility and time to individually follow up on older adults. This can ensure sufficient allocation of resources in CC when using and introducing eHealth. As previously mentioned, interoperability is still an issue. Policy makers and practices should continuously focus on ensuring access for all health care professionals to common CC eHealth solutions such as EHRs or patient portals.

Strengths and Limitations

In this scoping review, several limitations need to be addressed. First, we did not critically appraise the included studies as scoping reviews are flexible in their methodology [32] and are not always conducted when the aim is to map evidence [78]. By not critically appraising the studies, we included studies that varied greatly in the number of participants and methodological approaches. Therefore, our results should not be generalized. The results can be seen as important for future research and practice, policy makers, and the development of new eHealth tools to coordinate care for older adults. Furthermore, scoping reviews can include gray literature such as policies or government documents [32]. We excluded gray literature from this review as we aimed to map the research literature given the strong policy push to use eHealth [79].

Second, we searched 4 databases and used several search terms relevant to CC, eHealth, home care, and older adults. Our searches were conducted with the assistance of an experienced librarian. Despite our efforts to map the research literature on eHealth and CC for older adults living at home, we may have missed some studies.

Third, we included studies with participants aged ≥65 years and excluded several studies because the participants were younger (eg, aged 60 years). The World Health Organization [3] is moving away from using a chronological definition of old age (eg, 65 years). However, our decision was based on the need for knowledge on the use of eHealth in CC among older adults living at home [30].

Conclusions

The number of older adults will continue to increase well into the future. Older adults with chronic illnesses must navigate fragmented health care services, and eHealth in CC may be a way to prevent this fragmentation. The use of eHealth in CC for older adults is promising, although the outcomes so far have been mixed. eHealth in CC may improve older adults' mental and physical health and reduce hospital admissions and readmissions. A barrier was hampered communication because of the lack of interoperability of the EHRs and patient portals, which seems to be an ongoing issue worldwide.

To ensure the successful use of eHealth in CC for older adults living at home, the eHealth used needs to be customized to each individual's care needs. Education and patient support should be individualized. The telephone is still important for some older adults, and future research and practice should consider using the telephone or in-person visits to close the digital divide. However, it is essential to ensure that older adults interested in and capable of using HIT can be offered eHealth in CC. This calls for individualized eHealth-enabled health care services for older adults. eHealth in CC has an immense potential for the future organization and development of health care services. Thus, more in-depth knowledge of eHealth at the crossroads of CC for older adults living at home is needed.

Acknowledgments

Funding was provided by the University of Stavanger, Norway.

Authors' Contributions

All authors contributed to designing and developing the study topic, research questions, and literature search. HMHF conducted the literature search and charted the data with input from AMLH and MS. The data analysis was led by HMHF in collaboration with AMLH and MS. All authors contributed to the screening of articles and reading of the included articles, which was led by HMHF. HMHF was responsible for drafting the main text. All authors took part in reading, providing input, reviewing and editing the text, and approving the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of a search.
[DOCX File , 16 KB - jmir_v24i10e39584_app1.docx]

Multimedia Appendix 2
Overview of the included papers and care coordination activities.
[DOCX File , 25 KB - jmir_v24i10e39584_app2.docx]

References

1. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev 2011 Sep;10(4):430-439. [doi: 10.1016/j.arr.2011.03.003] [Medline: 21402176]
2. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
3. World report on ageing and health. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/186463 [accessed 2022-02-12]
4. Onder G, Palmer K, Navickas R, Jurevičienė E, Mammarella F, Strandzheva M, Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle (JA-CHRODIS). Time to face the challenge of multimorbidity. A European perspective from the joint action on chronic diseases and promoting healthy ageing across the life cycle (JA-CHRODIS). Eur J Intern Med 2015 Apr;26(3):157-159. [doi: 10.1016/j.ejim.2015.02.020] [Medline: 25797840]
5. Weiss CO. Frailty and chronic diseases in older adults. Clin Geriatr Med 2011 Feb;27(1):39-52. [doi: 10.1016/j.cger.2010.08.003] [Medline: 21093721]
6. Kämpfen F, Wijemunige N, Evangelista Jr B. Aging, non-communicable diseases, and old-age disability in low- and middle-income countries: a challenge for global health. Int J Public Health 2018 Dec;63(9):1011-1012. [doi: 10.1007/s00038-018-1137-z] [Medline: 29951746]

7. Global action plan on physical activity 2018-2030: More active people for a healthier world. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf [accessed 2022-02-12]
8. Naylor MD, Kurtzman ET, Pauly MV. Transitions of elders between long-term care and hospitals. Policy Polit Nurs Pract 2009 Aug;10(3):187-194. [doi: 10.1177/1527154409355710] [Medline: 20026453]
9. Elliott J, Stolee P, Boscart V, Giangregorio L, Heckman G. Coordinating care for older adults in primary care settings: understanding the current context. BMC Fam Pract 2018 Aug 07;19(1):137 [FREE Full text] [doi: 10.1186/s12875-018-0821-7] [Medline: 30086707]
10. McGilton KS, Vellani S, Yeung L, Chishtie J, Commisso E, Ploeg J, et al. Identifying and understanding the health and social care needs of older adults with multiple chronic conditions and their caregivers: a scoping review. BMC Geriatr 2018 Oct 01;18(1):231 [FREE Full text] [doi: 10.1186/s12877-018-0925-x] [Medline: 30285641]
11. McDonald KM, Schultz E, Albin L, Pineda N, Lonhart J, Sundaram V, et al. Care coordination measures atlas version 4. Agency for Healthcare Research and Quality. Rockville, MD, USA: Agency for Healthcare Research and Quality; 2014 Jun. URL: https://www.ahrq.gov/sites/default/files/publications/files/ccm_atlas.pdf [accessed 2022-02-12]
12. McDonald K, Sundaram V, Bravata DM, Lewis R, Lin N, Kraft SA, et al. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 7: Care Coordination). Rockville, MD, USA: Agency for Healthcare Research and Quality; 2007.
13. Transforming and scaling up health professionals' education and training: World Health Organization guidelines 2013. World Health Organization. 2019 Oct 1. URL: https://www.who.int/publications/i/item/transforming-and-scaling-up-health-professionals%E2%80%99-education-and-training [accessed 2022-02-12]
14. Dixon BE, Embi PJ, Haggstrom DA. Information technologies that facilitate care coordination: provider and patient perspectives. Transl Behav Med 2018 May 23;8(3):522-525. [doi: 10.1093/tbm/ibx086] [Medline: 29800411]
15. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/252529 [accessed 2022-02-12]
16. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
17. Kastner M, Cardoso R, Lai Y, Treister V, Hamid JS, Hayden L, et al. Effectiveness of interventions for managing multiple high-burden chronic diseases in older adults: a systematic review and meta-analysis. CMAJ 2018 Aug 27;190(34):E1004-E1012 [FREE Full text] [doi: 10.1503/cmaj.171391] [Medline: 30150242]
18. Abdi S, Spann A, Borilovic J, de Witte L, Hawley M. Understanding the care and support needs of older people: a scoping review and categorisation using the WHO international classification of functioning, disability and health framework (ICF). BMC Geriatr 2019 Jul 22;19(1):195 [FREE Full text] [doi: 10.1186/s12877-019-1189-9] [Medline: 31331279]
19. Dostálová V, Bártová A, Bláhová H, Holmerová I. The needs of older people receiving home care: a scoping review. Aging Clin Exp Res 2021 Mar;33(3):495-504. [doi: 10.1007/s40520-020-01505-3] [Medline: 32086715]
20. Feldman S, Dickins ML, Browning CJ, DeSoysa TS. The health and service needs of older veterans: a qualitative analysis. Health Expect 2015 Dec;18(6):2202-2212 [FREE Full text] [doi: 10.1111/hex.12190] [Medline: 24661337]
21. Breitholtz A, Snellman I, Fagerberg I. Living with uncertainty: older persons' lived experience of making independent decisions over time. Nurs Res Pract 2013;2013:403717 [FREE Full text] [doi: 10.1155/2013/403717] [Medline: 23533743]
22. Bloom DE, Chatterji S, Kowal P, Lloyd-Sherlock P, McKee M, Rechel B, et al. Macroeconomic implications of population ageing and selected policy responses. Lancet 2015 Feb 14;385(9968):649-657 [FREE Full text] [doi: 10.1016/S0140-6736(14)61464-1] [Medline: 25468167]
23. Kok L, Berden C, Sadiraj K. Costs and benefits of home care for the elderly versus residential care: a comparison using propensity scores. Eur J Health Econ 2015 Mar;16(2):119-131. [doi: 10.1007/s10198-013-0557-1] [Medline: 24390144]
24. Amalberti R, Nicklin W, Braithwaite J. Preparing national health systems to cope with the impending tsunami of ageing and its associated complexities: towards more sustainable health care. Int J Qual Health Care 2016 Jun;28(3):412-414. [doi: 10.1093/intqhc/mzw021] [Medline: 26980115]
25. van de Pol MH, Fluit CR, Lagro J, Niessen D, Rikkert MG, Lagro-Janssen AL. Quality care provision for older people: an interview study with patients and primary healthcare professionals. Br J Gen Pract 2015 Aug;65(637):e500-e507 [FREE Full text] [doi: 10.3399/bjgp15X686065] [Medline: 26212845]
26. Husebø AM, Storm M. Virtual visits in home health care for older adults. ScientificWorldJournal 2014;2014:689873 [FREE Full text] [doi: 10.1155/2014/689873] [Medline: 25506616]
27. Fryer AK, Friedberg MW, Thompson RW, Singer SJ. patient perceptions of integrated care and their relationship to utilization of emergency, inpatient and outpatient services. Healthc (Amst) 2017 Dec;5(4):183-193. [doi: 10.1016/j.hjdsi.2016.12.005] [Medline: 28117243]
28. Kooij L, Groen WG, van Harten WH. The effectiveness of information technology-supported shared care for patients with chronic disease: a systematic review. J Med Internet Res 2017 Jun 22;19(6):e221 [FREE Full text] [doi: 10.2196/jmir.7405] [Medline: 28642218]

29. Peterson K, Anderson J, Bourne D, Charns MP, Gorin SS, Hynes DM, et al. Health care coordination theoretical frameworks: a systematic scoping review to increase their understanding and use in practice. J Gen Intern Med 2019 May;34(Suppl 1):90-98 [FREE Full text] [doi: 10.1007/s11606-019-04966-z] [Medline: 31098976]
30. MacInnes J, Billings J, Dima AL, Farmer C, Nijpels G. Achieving person-centredness through technologies supporting integrated care for older people living at home: an integrative review. J Integr Care 2021 Jun 10;29(3):274-294. [doi: 10.1108/jica-03-2021-0013]
31. 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]
32. 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]
33. 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]
34. Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4: software for research synthesis. EPPI-Centre Software. London, UK: Social Science Research Unit, UCL Institute of Education; 2010. URL: https://eppi.ioe.ac.uk/cms/er4/Features/tabid/3396/Default.aspx [accessed 2022-02-12]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. 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]
37. Makai P, Perry M, Robben SH, Schers H, Heinen M, Olde Rikkert MG, et al. Which frail older patients use online health communities and why? A mixed methods process evaluation of use of the Health and Welfare portal. J Med Internet Res 2014 Dec 17;16(12):e278 [FREE Full text] [doi: 10.2196/jmir.3609] [Medline: 25519769]
38. Makai P, Perry M, Robben SH, Schers HJ, Heinen MM, Olde Rikkert MG, et al. Evaluation of an eHealth intervention in chronic care for frail older people: why adherence is the first target. J Med Internet Res 2014 Jun 23;16(6):e156 [FREE Full text] [doi: 10.2196/jmir.3057] [Medline: 24966146]
39. Gellis ZD, Kenaley B, McGinty J, Bardelli E, Davitt J, Ten Have T. Outcomes of a telehealth intervention for homebound older adults with heart or chronic respiratory failure: a randomized controlled trial. Gerontologist 2012 Aug;52(4):541-552. [doi: 10.1093/geront/gnr134] [Medline: 22241810]
40. Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. J Am Geriatr Soc 2014 May;62(5):889-895. [doi: 10.1111/jgs.12776] [Medline: 24655228]
41. Gokalp H, de Folter J, Verma V, Fursse J, Jones R, Clarke M. Integrated telehealth and telecare for monitoring frail elderly with chronic disease. Telemed J E Health 2018 Dec;24(12):940-957 [FREE Full text] [doi: 10.1089/tmj.2017.0322] [Medline: 30129884]
42. Sheeran T, Rabinowitz T, Lotterman J, Reilly CF, Brown S, Donehower P, et al. Feasibility and impact of telemonitor-based depression care management for geriatric homecare patients. Telemed J E Health 2011 Oct;17(8):620-626 [FREE Full text] [doi: 10.1089/tmj.2011.0011] [Medline: 21780942]
43. Logue MD, Effken JA. An exploratory study of the personal health records adoption model in the older adult with chronic illness. Inform Prim Care 2012;20(3):151-169 [FREE Full text] [doi: 10.14236/jhi.v20i3.21] [Medline: 23710840]
44. Lewis C, Moore Z, Doyle F, Martin A, Patton D, Nugent LE. A community virtual ward model to support older persons with complex health care and social care needs. Clin Interv Aging 2017 Jun 26;12:985-993 [FREE Full text] [doi: 10.2147/CIA.S130876] [Medline: 28721026]
45. de Jong CC, Ros WJ, van Leeuwen M, Witkamp L, Schrijvers G. Professionals' use of a multidisciplinary communication tool for patients with dementia in primary care. Comput Inform Nurs 2018 Apr;36(4):193-198. [doi: 10.1097/CIN.0000000000000414] [Medline: 29360700]
46. Biese K, Lamantia M, Shofer F, McCall B, Roberts E, Stearns SC, et al. A randomized trial exploring the effect of a telephone call follow-up on care plan compliance among older adults discharged home from the emergency department. Acad Emerg Med 2014 Feb;21(2):188-195 [FREE Full text] [doi: 10.1111/acem.12308] [Medline: 24673675]
47. Mavandadi S, Benson A, DiFilippo S, Streim JE, Oslin D. A telephone-based program to provide symptom monitoring alone vs symptom monitoring plus care management for late-life depression and anxiety: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1211-1218. [doi: 10.1001/jamapsychiatry.2015.2157] [Medline: 26558530]
48. Gurwitz JH, Field TS, Ogarek J, Tjia J, Cutrona SL, Harrold LR, et al. An electronic health record-based intervention to increase follow-up office visits and decrease rehospitalization in older adults. J Am Geriatr Soc 2014 May;62(5):865-871 [FREE Full text] [doi: 10.1111/jgs.12798] [Medline: 24779524]

49. Cutrona SL, Fouayzi H, Burns L, Sadasivam RS, Mazor KM, Gurwitz JH, et al. Primary care providers' opening of time-sensitive alerts sent to commercial electronic health record InBaskets. J Gen Intern Med 2017 Nov;32(11):1210-1219 [FREE Full text] [doi: 10.1007/s11606-017-4146-3] [Medline: 28808942]
50. Mateo-Abad M, Fullaondo A, Merino M, Gris S, Marchet F, Avolio F, CareWell project group. Impact assessment of an innovative integrated care model for older complex patients with multimorbidity: the CareWell project. Int J Integr Care 2020 May 22;20(2):8 [FREE Full text] [doi: 10.5334/ijic.4711] [Medline: 32477037]
51. Dent M, Tutt D. Electronic patient information systems and care pathways: the organisational challenges of implementation and integration. Health Informatics J 2014 Sep;20(3):176-188 [FREE Full text] [doi: 10.1177/1460458213518545] [Medline: 24550562]
52. Freilich J, Nilsson GH, Ekstedt M, Flink M. "Standing on common ground" - a qualitative study of self-management support for patients with multimorbidity in primary health care. BMC Fam Pract 2020 Nov 17;21(1):233 [FREE Full text] [doi: 10.1186/s12875-020-01290-y] [Medline: 33203401]
53. Hawley CE, Genovese N, Owsiany MT, Triantafylidis LK, Moo LR, Linsky AM, et al. Rapid integration of home telehealth visits amidst COVID-19: what do older adults need to succeed? J Am Geriatr Soc 2020 Nov;68(11):2431-2439 [FREE Full text] [doi: 10.1111/jgs.16845] [Medline: 32930391]
54. Chu JN, Kaplan C, Lee JS, Livaudais-Toman J, Karliner L. Increasing telehealth access to care for older adults during the COVID-19 pandemic at an academic medical center: Video Visits for Elders Project (VVEP). Jt Comm J Qual Patient Saf 2022 Mar;48(3):173-179 [FREE Full text] [doi: 10.1016/j.jcjq.2021.11.006] [Medline: 35027304]
55. Melchiorre MG, Lamura G, Barbabella F, ICARE4EU Consortium. eHealth for people with multimorbidity: results from the ICARE4EU project and insights from the "10 e's" by Gunther Eysenbach. PLoS One 2018 Nov 14;13(11):e0207292 [FREE Full text] [doi: 10.1371/journal.pone.0207292] [Medline: 30427924]
56. Schultz EM, Pineda N, Lonhart J, Davies SM, McDonald KM. A systematic review of the care coordination measurement landscape. BMC Health Serv Res 2013 Mar 28;13:119 [FREE Full text] [doi: 10.1186/1472-6963-13-119] [Medline: 23537350]
57. Garnett D, Hardy L, Fitzgerald E, Fisher T, Graham L, Overcash J. Nurse case manager: measurement of care coordination activities and quality and resource use outcomes when caring for the complex patient with hematologic cancer. Clin J Oncol Nurs 2020 Feb 01;24(1):65-74. [doi: 10.1188/20.CJON.65-74] [Medline: 31961839]
58. Chakurian D, Popejoy L. Utilizing the care coordination Atlas as a framework: an integrative review of transitional care models. Int J Care Coord 2021 Mar 12;24(2):57-71. [doi: 10.1177/20534345211001615]
59. Samal L, Dykes PC, Greenberg JO, Hasan O, Venkatesh AK, Volk LA, et al. Care coordination gaps due to lack of interoperability in the United States: a qualitative study and literature review. BMC Health Serv Res 2016 Apr 22;16:143 [FREE Full text] [doi: 10.1186/s12913-016-1373-y] [Medline: 27106509]
60. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 09;288(14):1775-1779. [doi: 10.1001/jama.288.14.1775] [Medline: 12365965]
61. Tith D, Lee JS, Suzuki H, Wijesundara WM, Taira N, Obi T, et al. Application of blockchain to maintaining patient records in electronic health record for enhanced privacy, scalability, and availability. Healthc Inform Res 2020 Jan;26(1):3-12 [FREE Full text] [doi: 10.4258/hir.2020.26.1.3] [Medline: 32082695]
62. Regulation GDPR EU 2016/679 of the European Parliament and of the Council of 27 April 2016. Official Journal of the European Union. 2016. URL: https://www.eumonitor.eu/9353000/1/j4nvk6yhcbpeywk_j9vvik7m1c3gyxp/vk3t7p3lbczq#:~:text=Regulation%20(EU)%202016%2F679%20of%20the%20European%20Parliament%20and,(OJ%20L%20119%2C%204.5 [accessed 2022-02-12]
63. Anderson M, Perrin A. Technology use among seniors. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2022-02-12]
64. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
65. Ali MA, Alam K, Taylor B, Ashraf M. Examining the determinants of eHealth usage among elderly people with disability: the moderating role of behavioural aspects. Int J Med Inform 2021 May;149:104411. [doi: 10.1016/j.ijmedinf.2021.104411] [Medline: 33618191]
66. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. Psychol Aging 2013 Dec;28(4):1124-1139 [FREE Full text] [doi: 10.1037/a0033911] [Medline: 24364414]
67. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
68. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
69. Rios G. eHealth literacy and older adults: a review of literature. Top Geriatr Rehabil 2013;29(2):115-125. [doi: 10.1097/tgr.0b013e31827ec0bd]
70. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Info Systems J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]

71. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
72. Vroman KG, Arthanat S, Lysack C. "Who over 65 is online?" Older adults' dispositions toward information communication technology. Comput Human Behav 2015 Feb;43:156-166. [doi: 10.1016/j.chb.2014.10.018]
73. Terp R, Kayser L, Lindhardt T. Older patients' competence, preferences, and attitudes toward digital technology use: explorative study. JMIR Hum Factors 2021 May 14;8(2):e27005 [FREE Full text] [doi: 10.2196/27005] [Medline: 33988512]
74. Gill E, Dykes PC, Rudin RS, Storm M, McGrath K, Bates DW. Technology-facilitated care coordination in rural areas: what is needed? Int J Med Inform 2020 May;137:104102 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104102] [Medline: 32179256]
75. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
76. Liaw ST, Kearns R, Taggart J, Frank O, Lane R, Tam M, et al. The informatics capability maturity of integrated primary care centres in Australia. Int J Med Inform 2017 Sep;105:89-97. [doi: 10.1016/j.ijmedinf.2017.06.002] [Medline: 28750915]
77. Hsiao CJ, King J, Hing E, Simon AE. The role of health information technology in care coordination in the United States. Med Care 2015 Feb;53(2):184-190 [FREE Full text] [doi: 10.1097/MLR.0000000000000276] [Medline: 25464164]
78. 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]
79. Asthana S, Jones R, Sheaff R. Why does the NHS struggle to adopt eHealth innovations? A review of macro, meso and micro factors. BMC Health Serv Res 2019 Dec 21;19(1):984 [FREE Full text] [doi: 10.1186/s12913-019-4790-x] [Medline: 31864370]

Abbreviations

BP: blood pressure
CC: care coordination
ED: emergency department
EHR: electronic health record
HIT: health IT
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
RCT: randomized controlled trial

Edited by T Leung; submitted 19.05.22; peer-reviewed by E Gill, N Ivers, S de Sequeira; comments to author 13.07.22; revised version received 09.08.22; accepted 10.08.22; published 18.10.22.

Please cite as:
Fjellså HMH, Husebø AML, Storm M
eHealth in Care Coordination for Older Adults Living at Home: Scoping Review
J Med Internet Res 2022;24(10):e39584
URL: https://www.jmir.org/2022/10/e39584
doi:10.2196/39584
PMID:36256831

Review

Understanding How the Design and Implementation of Online Consultations Affect Primary Care Quality: Systematic Review of Evidence With Recommendations for Designers, Providers, and Researchers

Sarah Darley[1], PhD; Tessa Coulson[2], MRes; Niels Peek[3,4], PhD; Susan Moschogianis[1], PhD; Sabine N van der Veer[3,5], PhD; David C Wong[3,6], DPhil; Benjamin C Brown[1,3,4], PhD

[1]Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester, Manchester, United Kingdom

[2]National Health Service Salford Clinical Commissioning Group, Salford, United Kingdom

[3]Centre for Health Informatics, Division of Informatics, Imaging and Data Science, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom

[4]National Institute for Health and Care Research Greater Manchester Patient Safety Translational Research Centre, School of Health Sciences, The University of Manchester, Manchester, United Kingdom

[5]National Institute for Health and Care Research Applied Research Collaboration Greater Manchester, Manchester, United Kingdom

[6]Department of Computer Science, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Benjamin C Brown, PhD
Division of Population Health, Health Services Research and Primary Care
School of Health Sciences
The University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 7919103175
Email: Benjamin.brown@manchester.ac.uk

Abstract

Background: Online consultations (OCs) allow patients to contact their care providers on the web. Worldwide, OCs have been rolled out in primary care rapidly owing to policy initiatives and COVID-19. There is a lack of evidence regarding how OC design and implementation influence care quality.

Objective: We aimed to synthesize research on the impacts of OCs on primary care quality, and how these are influenced by system design and implementation.

Methods: We searched databases from January 2010 to February 2022. We included quantitative and qualitative studies of real-world OC use in primary care. Quantitative data were transformed into qualitative themes. We used thematic synthesis informed by the Institute of Medicine domains of health care quality, and framework analysis informed by the nonadoption, abandonment, scale-up, spread, and sustainability framework. Strength of evidence was judged using the GRADE-CERQual approach.

Results: We synthesized 63 studies from 9 countries covering 31 OC systems, 14 (22%) of which used artificial intelligence; 41% (26/63) of studies were published from 2020 onward, and 17% (11/63) were published after the COVID-19 pandemic. There was no quantitative evidence for negative impacts of OCs on patient safety, and qualitative studies suggested varied perceptions of their safety. Some participants believed OCs improved safety, particularly when patients could describe their queries using free text. Staff workload decreased when sufficient resources were allocated to implement OCs and patients used them for simple problems or could describe their queries using free text. Staff workload increased when OCs were not integrated with other software or organizational workflows and patients used them for complex queries. OC systems that required patients to describe their queries using multiple-choice questionnaires increased workload for patients and staff. Health costs decreased when patients used OCs for simple queries and increased when patients used them for complex queries. Patients using OCs were more likely

to be female, younger, and native speakers, with higher socioeconomic status. OCs increased primary care access for patients with mental health conditions, verbal communication difficulties, and barriers to attending in-person appointments. Access also increased by providing a timely response to patients' queries. Patient satisfaction increased when using OCs owing to better primary care access, although it decreased when using multiple-choice questionnaire formats.

Conclusions: This is the first theoretically informed synthesis of research on OCs in primary care and includes studies conducted during the COVID-19 pandemic. It contributes new knowledge that, in addition to having positive impacts on care quality such as increased access, OCs also have negative impacts such as increased workload. Negative impacts can be mitigated through appropriate OC system design (eg, free text format), incorporation of advanced technologies (eg, artificial intelligence), and integration into technical infrastructure (eg, software) and organizational workflows (eg, timely responses).

Trial Registration: PROSPERO CRD42020191802; https://tinyurl.com/2p84ezjy

(J Med Internet Res 2022;24(10):e37436) doi:10.2196/37436

KEYWORDS

general practice; systematic review; remote consultation; OC; triage; primary health care; care provider; health care professional; workforce; telemedicine; COVID-19; pandemic; primary care; health outcome; patient care

Introduction

Background

Online consultation (OC) systems allow patients to contact their health care provider over the internet to ask health-related questions and report symptoms [1]. Their query may then be resolved with a written response, telephone call, video consultation, or in-person visit. Many terms are used to describe this type of technology, including *e-consultation*, *e-visit*, and *online triage* (Multimedia Appendix 1 [2-28])—in this review, we refer to them all as *online consultations*. We distinguish OCs from "symptom checkers" [29] and other self-service systems that typically do not directly facilitate communication with a human health care provider and from patient portals [30], which may include generic email or secure messaging functionalities.

OCs are considered by policy makers in many countries as a way to address the increasing workload and decreasing workforce capacity in primary care [31-36] while still meeting patient expectations and improving access [37]. However, they have the potential to exacerbate health inequities [38,39] and increase inappropriate antibiotic prescriptions [40]. Furthermore, there are widely recognized challenges in initiating and sustaining the adoption of new technologies in primary care [41].

Although symptom checkers [29,42] and patient portals [30,43,44] have been well studied, only a small number of evidence syntheses directly relevant to OCs have been published: a systematic review of 57 articles on delivering "e-consultation" in primary care largely focused on generic stand-alone applications such as email and video (n=39/57, 68%) [45]; a scoping review of "online triage tools" included 13 papers, 4 of which (31%) were nonempirical (eg, opinion pieces) [46]; and a review of 17 studies of "intelligent online triage tools" focused only on those that used "artificial intelligence" (AI) [47].

Since these syntheses were conducted, OCs have gained wider traction in clinical practice worldwide—they have been indispensable in helping manage patients remotely to minimize the spread of COVID-19 [48,49], and English primary care providers have been mandated to offer OCs for all patients since April 2020 [50]. Moreover, OC system product design has progressed significantly to become more specialized and technologically advanced [51], with several more empirical research studies published on their use [2-11,52-64].

Given this rapid scale-up and increase in the diversity and complexity of OCs, further insight is needed into their impact on health care quality. Previous reviews have not reported the design or implementation details of the OCs they studied [45-47] despite their importance in understanding the causal mechanisms of how they affect care outcomes [65]. The aim of this study was to systematically review and synthesize the empirical quantitative and qualitative literature in a theoretically informed way to address this knowledge gap.

Objectives

Informed by existing theories, the aim of this study was to synthesize quantitative and qualitative research on (1) the impacts of OCs on primary care quality and (2) how these are influenced by OC system design and implementation.

Methods

Study Design

We consider OCs as complex interventions and, therefore, synthesized both quantitative and qualitative evidence to understand their impacts in specific contexts [66]. We did not perform a meta-analysis because of the heterogeneous and nonrandomized nature of the included studies [67]. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [68].

Registration and Protocol

The study protocol was registered with PROSPERO (CRD42020191802) [69]. The original title was amended to be less general and more specific to the objectives of the review, and the objectives were amended to focus on care quality.

Inclusion Criteria

Papers that met the following criteria were included: empirical studies using quantitative or qualitative methods to examine the

real-world use of OCs in primary care in any country, written in English, and published in 2010 or later. We excluded news articles, opinion pieces, literature reviews, non–English-language articles, and literature published before 2010.

We defined OCs as digital interventions that allow patients to contact their primary care provider by inputting "queries" into health care–specific web-based forms [1]. We included symptom checkers and similar self-service systems [54] if at least one of their outcomes directly facilitated contact with a primary care health professional. We included patient portals if they had a secure messaging functionality that used health care–specific forms [54]. We excluded stand-alone generic communication technologies such as email or videoconferencing software.

Search Strategy

We searched the Ovid MEDLINE, EMBASE, Web of Science, and Scopus databases during July 2020 (Multimedia Appendix 2 [12,53,56,58-60,63,70-73]). Our search strategy was developed from scoping searches of the literature and drew on search strategies used in related literature reviews [45,46]. We searched the National Technical Information Service, the Health Management Information Consortium, and Zetoc to find relevant gray literature, conference proceedings, and theses. We found further literature through citation mapping and in the reference lists of the included papers, searching during August 2020 and September 2020. SD and TC independently screened titles and abstracts and then full papers for eligibility, resolving differences through discussion at each stage. All literature searches were rerun by SD between November 2021 and February 2022.

Data Extraction and Quality Appraisal

We extracted data from the included papers as verbatim text, capturing study characteristics (eg, research design and study setting) and key findings relevant to our research objectives based on the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [74] (Multimedia Appendix 3). We used the NASSS to capture "a rich, contextualised narrative of technology-supported change efforts and the numerous interacting influences that help explain its successes, failures, and unexpected events" [75]. The methodological quality of the studies was assessed using the Mixed Methods Appraisal Tool (MMAT), which is designed for qualitative, quantitative, and mixed methods studies [76]. We scored each paper using recommended quintile percentages as cutoffs and considered any paper scoring at least 60% as of "good" quality [77]. SD and TC extracted data from 10 papers independently, which confirmed high interrater agreement. Following this, SD extracted data from the remaining papers, which were checked by TC.

Data Synthesis

The data were imported into NVivo (version 12; QSR International) [78] for synthesis. To integrate both quantitative and qualitative data, during data synthesis, quantitative data were transformed into qualitative themes ("qualitising") [79].

For objective 1, we considered "impacts of OCs on primary care quality" as consequences of using OCs that could relate to patients, primary care staff, or the wider system [65]. We used thematic synthesis [80], which involved SD and TC coding the text from the data extraction forms independently line by line, developing higher-level themes through regular discussion [80]. Impacts on care quality were synthesized inductively, with emerging themes mapped to the six Institute of Medicine domains of health care quality [81]: safe (avoiding harm to patients from care that is intended to help), effective (providing care based on scientific knowledge to produce better clinical outcomes), patient-centered (care that is respectful and responsive), timely (reducing waits and delays for those who receive and give care), efficient (avoiding waste), and equitable (care that does not vary in quality because of personal characteristics) [81]. Our emergent findings suggested that OCs had both positive and negative impacts and, therefore, theme descriptions were edited to be neutral (eg, safe→safety and efficient→efficiency).

For objective 2, we considered OC "design" as material properties of an OC, such as features and functionality [74], and "implementation" as the way an OC was introduced and used in a particular context [65]. As a design feature, we considered AI as the ability of machines to "mimic human intelligence as characterized by behaviors such as cognitive ability, memory, learning, and decision making" [82]. We synthesized the extracted data using framework analysis [83], which involved SD and TC reading and rereading each data extraction form and then coding them line by line independently—both deductively by using domains from the NASSS framework [74] for high-level themes and inductively by identifying additional subthemes. Through discussion, SD and TC summarized the findings into five high-level themes: condition complexity (health condition and the illness the OC is used for), technology (material properties of the OC and required knowledge for use), adopters (staff, patients, and carers expected to use the OC), organization (extent of work needed for implementation of the OC, capacity, and readiness), and wider system (policy context) [74]. Two NASSS domains—value proposition (value of the OC to the developer, patients, and health care system) and embedding and adaptation over time (learning and adaptation to changing contexts)—had limited applicability to our findings and were not included in the final synthesis. Informed by realistic evaluation [65], we considered our themes as contextual factors and identified patterns of explanations for how each led to the impacts on care quality from objective 1 (ie, "causal mechanisms"). Where appropriate, we considered the levels of OC adoption as a mechanism for how they affected care quality [65]. We used visual mapping to identify commonalities and discordances in causal mechanisms—first within individual papers and then across papers [83]. Where there were discordances, we explored potential explanations where possible (eg, related to the study setting).

The strength and quality of our findings for objectives 1 and 2 were assessed using the Grading of Recommendations Assessment, Development, and Evaluation-Confidence in Evidence from Reviews of Qualitative Research method [84]. This accounts for the methodological limitations of the contributing papers (according to MMAT assessments),

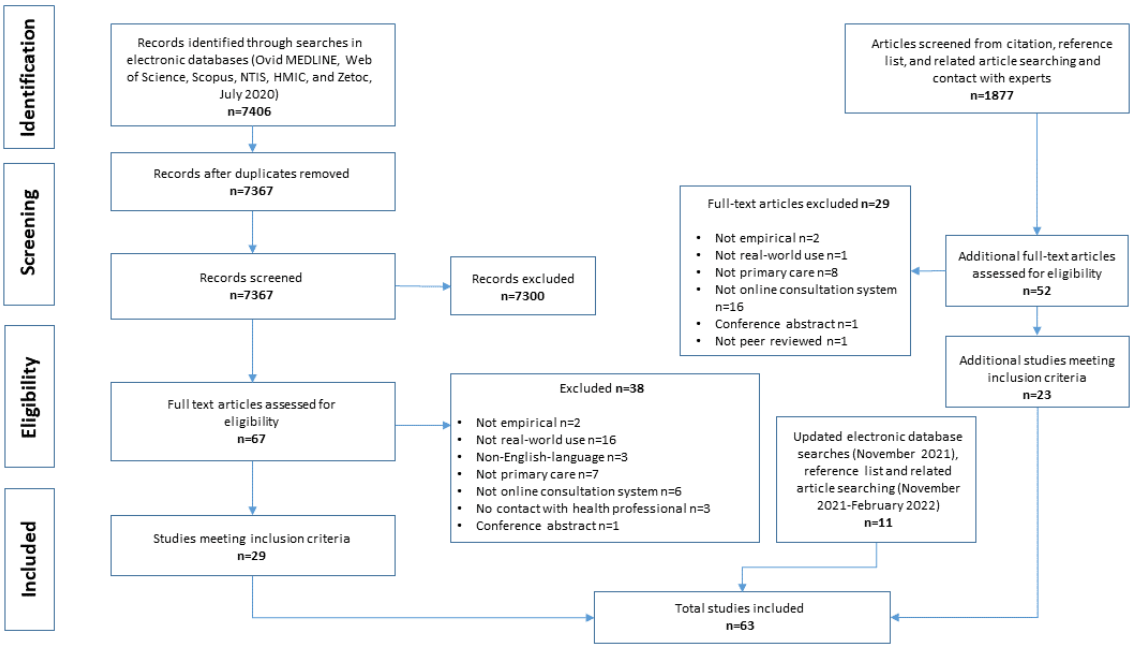
relevance to the review question, coherence of the finding, and adequacy of its supporting data [84]. Confidence in each finding was designated as high, moderate, low, or very low. At each stage of the analysis, the findings were discussed and agreed upon with the wider study team. BCB reviewed all coded verbatim excerpts from the papers included in the final synthesis.

Results

Descriptive Summary

We synthesized 63 papers (Figure 1), including 52 (83%) journal papers [53], 7 (11%) evaluation reports [85], 3 (5%) conference papers [12], and 1 (2%) master's degree thesis [13]. The studies were quantitative (33/63, 52%), qualitative (12/63, 19%), and mixed methods (18/63, 29%) and analyzed data from patients (16/63, 25% qualitative studies and 18/63, 29% quantitative studies), staff (22/63, 35% qualitative studies and 9/63, 14% quantitative studies), and clinical systems (33/63, 52% quantitative studies). All were set in one of 9 high-income countries, with most coming from the United States (21/63, 33%) and the United Kingdom (20/63, 32%; Multimedia Appendix 4 [2-27,52-64,70-73,77,85-104]). In all, 41% (26/63) of the studies were published in 2020 or later, and 17% (11/63) were conducted after the start of the COVID-19 pandemic. Examples of excluded studies are those that focused on stand-alone video consultations [105], involved communication between physicians and not patients [106], and were not based on primary care [107].

Figure 1. Flowchart of the study selection process. HMIC: Health Management Information Consortium; NTIS: National Technical Information Service.



In all, 83% (52/63) of the studies reported levels of OC adoption by patients and staff, of which 62% (32/52; 32/63, 51% of all studies) were described as "low" by the study authors [86]. OCs were adopted at a high rate in 63% (33/52; 33/63, 52% of all) of the studies [87], including high rates of adoption by certain patient groups even when overall OC adoption in the study was low [14].

The included papers described 31 OC systems summarized in Table 1 and detailed in Multimedia Appendix 5 [2-27,52-64,70-73,85-104]. In 25% (16/63) of the papers, the OC system was described sufficiently to meet our inclusion criteria but not in enough detail to determine specific design features. Of the 31 OCs described, most (23/31, 74%) offered two-way written communication between patients and staff [88], with a few (4/31, 13%) also offering communication by video [52]. In all, 13% (4/31) did not provide functionalities for staff to reply to patients via the system (ie, one-way communication only [14]). In total, 35% (11/31) required patients to describe their queries solely via multiple-choice questionnaires (MCQs) [89] compared with 13% (4/31) that solely required patients to describe their queries using unstructured free text [56]. In all, 42% (13/31) had a hybrid approach of primarily using MCQs with the option for patients to enter additional free text [90]. No free text OCs offered optional MCQs. In all, 26% (8/31) of the OC systems were integrated with the electronic health record (EHR) [58], and 3% (1/31) allowed patients to schedule telephone or in-person appointments with health care professionals themselves [54].

In total, 54% (13/24) of MCQ-based OC systems exhibited three types of AI: (1) adapting questions they asked patients as they submitted their query in response to previous answers given (10/31, 32%) [91]; (2) prioritizing patient queries based on clinical urgency (4/31, 13%) [54]; and (3) signposting patients to an appropriate care provider based on their query, such as self-care, primary care, or emergency department (3/31, 10%) [8]. These were mostly powered by preprogrammed logic and "algorithms" (10/31, 32%) [54], with the exact AI methodology unclear in the remainder (3/31, 10%) [15].

The methodological quality of most studies (42/63, 67%) was "good" (ie, ≥60% according to the MMAT [77]; Multimedia Appendix 6 [2-27,52-64,70-73,76,85-104]). Common limitations included a lack of detail on whether the OC was administered as intended [92] and small sample sizes [3].

Table 1. Online consultation (OC) system features (N=31).

OC system feature and subcategory	Studies, n (%)[a]
Communication mode	
Two-way written communication between staff and patients	23 (74)
One-way written communication (staff cannot reply to patients)	4 (13)
Videoconferencing	4 (13)
Unclear	4 (13)
Patient query format	
Multiple-choice questionnaires only	11 (35)
Unstructured free text only	4 (13)
Multiple-choice questionnaires with optional free text	13 (42)
Unclear	3 (10)
Integration with other software	
Electronic health record	8 (26)
Appointment scheduling	1 (3)
No integration	23 (74)
Artificial intelligence function	
Adapting questions during query submission	10 (32)
Prioritizing patient queries based on clinical urgency	4 (13)
Signposting patients to the most appropriate care provider	3 (10)
No artificial intelligence	17 (55)
Artificial intelligence method	
Preprogrammed logic and algorithms	10 (32)
Unclear	3 (10)

[a]Count of OC systems described in detail (n=31). Categories may add up to >31 as OC systems may have more than one feature in a category.

Synthesis

Overview

To maintain readability, we present only moderate- and high-confidence findings and provide only 1 example reference per finding. Tables 2 and 3 provide all the references and specify whether the findings are qualitative or quantitative. Multimedia Appendix 7 [13,59,99] and Multimedia Appendix 8 [3, 5, 8-11, 13-17, 19-21, 25, 27, 54, 57, 59, 60, 63, 64, 71, 85, 90, 91, 95, 97, 100, 101] detail the low-confidence findings. Multimedia Appendix 9 [2-27,52,54-61,63,64,67,70-73,85-101] and Multimedia Appendix 10 [3-27, 52, 54-64, 70-73, 85-95, 97, 98, 100, 101] provide exemplar data.

Table 2. Impacts of online consultations (OCs) on primary care quality.

Theme	Subtheme

Safety (harm to patients)	• Decreased patient safety (qualitative) [2,3,5,7,10,13,17,18,23-25,55,61,63,85,90,94]
	  • Description: patient and staff perceptions that OCs worsened patient safety
	  • CERQual[a] rating: high
	• Neutral-increased patient safety (qualitative and quantitative) [3-5,9,11,13,14,16,18,21,54,55,57-59,62,63,70,71,88,89,92,93,95,96]
	  • Description: no quantitative evidence of negative impacts on patient safety, with clinician and patient perceptions that OCs improved patient safety
	  • CERQual rating: high

Effective (providing care based on scientific knowledge to produce better clinical outcomes)	• Reduced antibiotic prescribing rates (quantitative) [15,60,62,97]
	  • Description: fewer antibiotics prescribed when using OCs
	  • CERQual rating: moderate

Timeliness (reducing waits and delays)	• Increased access (qualitative and quantitative) [2-4,6,7,9,13-21,23-25,55-58,62-64,85,90,92,95]
	  • Description: easier and more convenient for patients to contact their primary care provider and quicker to communicate with a health professional
	  • CERQual rating: high

Efficiency (avoiding waste)	• Decreased workload (qualitative and quantitative) [3-5,9,11,13-21,23,54-58,60,61,63,64,70,71,85,89,90,92-95]
	  • Description: less work for staff and patients to provide and receive care, respectively
	  • CERQual rating: high
	• Increased workload (qualitative and quantitative) [3-5,8-10,13-23,25,52,55,56,58,64,85-87,92,93,98]
	  • Description: more work for staff and patients to provide and receive care, respectively
	  • CERQual rating: high
	• Decreased costs (qualitative and quantitative) [5,15-18,21,23,56,57,60,61,63,70,85,89,92,95,96,99,100]
	  • Description: lower costs for the health care system and patients to provide and receive care, respectively
	  • CERQual rating: high
	• Increased costs (qualitative and quantitative) [5,16-19,22,23,63,87]
	  • Description: higher costs for the health care system
	  • CERQual rating: high

Equitable (variation because of personal characteristics)	• Decreased equity (qualitative and quantitative) [7,8,12-27,52,57,59,60,63,64,70-73,85,87-92,94,95,97,98,100,101]
	  • Description: OC use variation based on patient characteristics
	  • CERQual rating: high
	• Increased equity (qualitative) [7,9,14-20,23,24,27,57,63,64,85,87,90,91]
	  • Description: OCs helped patients who had previously struggled because of their personal characteristics communicate with their primary care providers
	  • CERQual rating: high

Patient-centeredness (care that is respectful and responsive)	• Decreased patient satisfaction (qualitative) [9,11,14,15,18,21,23-25,57,64,85,90]
	  • Description: negative patient experiences of using OCs
	  • CERQual rating: high
	• Increased patient satisfaction (qualitative and quantitative) [2,5-7,9,11,13-21,23-25,56,57,63,64,71,85,89,90,92-94,96,99]
	  • Description: positive patient experiences of using OCs
	  • CERQual rating: high

[a]CERQual: Confidence in the Evidence from Reviews of Qualitative Research.

Table 3. How the impacts of online consultations (OCs) on primary care quality are influenced by system design and implementation.

Theme and OC design feature or implementation	Impact on care quality (from Table 2)[a]	CERQual[b] rating and references
Condition complexity (illness the OC is used for)		
• Decreased complexity of query • Description: patient queries are straightforward and easy to resolve (eg, administrative tasks, minor acute illnesses, and prescription requests)	• Efficiency: decreased workload (qualitative and quantitative) • Efficiency: decreased health costs (qualitative and quantitative)	• CERQual rating: high [5,15-18,23,56,61,64,70,85]
• Increased complexity of query • Description: patient queries are not straightforward and easy to resolve (eg, multiple ill-defined symptoms)	• Efficiency: increased workload (qualitative) • Efficiency: increased health costs (qualitative and quantitative)	• CERQual rating: high [5,16-19,22,23]
Technology (material properties of the OC)		
• MCQs[c] • Description: patients describe their query by completing questionnaires and selecting their answers from a list	• Efficiency: increased workload (qualitative) • Patient-centeredness: decreased patient satisfaction (qualitative)	• CERQual rating (efficiency): high [5, 9, 10, 14, 17, 18, 20, 21, 23, 25, 55, 64, 86] • CERQual rating (patient-centeredness): high [5,9,14,18,20,21,25,64,86]
• Free text input • Description: patients describe their query using unstructured text	• Efficiency: decreased workload (qualitative and quantitative) • Safety: increased patient safety (qualitative)	• CERQual rating: high [3,16,21,55,58,93,95]
• Two-way written communication • Description: patients and staff are able to send written messages to each other	• Efficiency: decreased workload (qualitative and quantitative)	• CERQual rating: high [55-58,94,95]
• Nonintegration with core software systems • Description: OC systems that operate separately from other software used by the primary care provider	• Efficiency: increased workload (qualitative)	• CERQual rating: high [3-5,10,13,15,17-21,23,55]
Adopters (expected users of OCs)		
• Female sex • Description: female patients	• High adoption (qualitative and quantitative) • Equitable: decreased equity (qualitative and quantitative)	• CERQual rating: high [8, 12, 13, 15, 18, 20-23, 27, 52, 57, 60, 70, 72, 73, 87-92, 94, 95, 97, 100, 101]
• Lower age • Description: younger patients	• High adoption (qualitative and quantitative) • Equitable: decreased equity (qualitative and quantitative)	• CERQual rating: high [7, 8, 13-15, 18, 19, 21-23, 27, 52, 59, 63, 64, 70, 71, 73, 85, 87-91, 94, 97, 101]
• Native speakers • Description: patients who are native speakers of the official language of the country they live in	• High adoption (qualitative and quantitative) • Equitable: decreased equity (qualitative and quantitative)	• CERQual rating: high [18,23,25,57,63,89,98]
• High socioeconomic status • Description: patients with higher levels of income and education	• High adoption (qualitative and quantitative) • Equitable: decreased equity (qualitative and quantitative)	• CERQual rating: high [15,18,23-27,57,85,87,90]
• Mental health conditions • Description: patients with a mental health diagnosis	• Timeliness: increased access (qualitative) • Equitable: increased equity (qualitative) • Patient-centeredness: increased patient satisfaction (qualitative and quantitative)	• CERQual rating: high [9,14,15,18-20,57,64]
• Verbal communication difficulties • Description: patients with difficulty communicating verbally (eg, those with hearing loss)	• Timeliness: increased access (qualitative) • Equitable: increased equity (qualitative) • Patient-centeredness: increased patient satisfaction (qualitative and quantitative)	• CERQual rating: high [16-19,24,64,90]

Theme and OC design feature or implementation	Impact on care quality (from Table 2)[a]	CERQual[b] rating and references
• Physical barriers to attending in-person appointments • Description: patients cannot easily attend in-person appointments (eg, because of physical disabilities, living far from their primary care provider, work commitments, or care responsibilities)	• Timeliness: increased access (qualitative) • Equitable: increased equity (qualitative) • Patient-centeredness: increased patient satisfaction (qualitative and quantitative)	• CERQual rating: high [7,15,18,20,23,63,64,85]
• Preference for traditional consulting methods • Description: staff and patients believe in-person consultations are the gold standard	• Low adoption (qualitative)	• CERQual rating: high [11,18,19,24,26,63,85,93]
Organization (work needed to implement OCs)		
• Lack of OC promotion • Description: patients are not effectively informed that OCs are available for them to contact their primary care provider	• Low adoption (qualitative and quantitative)	• CERQual rating: moderate [16,18,24,26,95]
• Timely response • Description: primary care providers respond quickly to patients' OC queries	• Patient-centeredness: increased patient satisfaction (qualitative and quantitative) • Timeliness: increased access (qualitative)	• CERQual rating: high [6,13,20,21,23,25,57]
• Nonintegration with daily workflows • Description: primary care provider does not coherently plan OCs into their work processes (eg, by not scheduling clinician time to deal with OCs or not diverting as much incoming patient demand as possible via OCs)	• Efficiency: increased workload (qualitative and quantitative)	• CERQual rating: high [4,5,13,14,17-20,52,55,85,86,93]
• Sufficient resources allocated to implementing OCs • Description: adequate training, staff, and facilities are available to conduct OCs	• Efficiency: decreased workload (qualitative)	• CERQual rating: high [5,13-15,55,85,86,93]
• Lack of continuity of care • Description: OC query is not dealt with by a known or preferred physician	• Patient-centeredness: decreased patient satisfaction (qualitative)	• CERQual rating: moderate [6,13,15,64,92]
Wider system (policy context)		
• Government policy • Description: policies mandating OC use (eg, by increasing digital modes of contact with primary care in general or minimizing in-person contact during the COVID-19 pandemic)	• High adoption (qualitative and quantitative)	• CERQual rating: high [4,15,54,62,63,87]
• Lack of financial support • Description: no external funding available to pay ongoing costs of OCs	• Low adoption (qualitative and quantitative)	• CERQual rating: moderate [5,18,23,63,85]

[a]Includes levels of OC adoption as a mechanism for how they affect care quality [65].

[b]CERQual: Confidence in the Evidence from Reviews of Qualitative Research.

[c]MCQ: multiple-choice questionnaire.

Objective 1: Impacts of OCs on Primary Care Quality

Safety

In 27% (17/63) of the studies, staff and patients expressed general concerns about the impact of OCs on patient safety, particularly regarding the potential loss of information from patients versus in-person or telephone consultations and how it could lead to misdiagnosis [55]. However, quantitative evidence from 17% (11/63) of the studies did not support these concerns in terms of emergency department attendance rates [92], hospitalizations [70], deaths [88], and other measures [59]. Furthermore, clinicians and patients in 22% (14/63) of the studies believed that OCs improved patient safety, for example, by producing a detailed shared written record of consultations [93] and helping reduce the spread of communicable diseases such as COVID-19 [63].

Effectiveness

In 6% (4/63) of the studies, antibiotics were prescribed to patients at a lower rate via OCs compared with in-person consultations [60].

Timeliness

In 46% (29/63) of the studies, OCs were perceived as increasing access to primary care services. It was easier and more convenient to make initial contact as patients could submit an OC query at any time without waiting on the phone or attending in person [14]. Once a query was submitted, patients also communicated with health professionals sooner as OCs tended to circumvent the traditional appointment-booking process [57].

Efficiency

In total, 52% (33/63) of the studies suggested that the workload decreased for both staff and patients when using OCs. Patient queries were written rather than spoken, incoming phone calls to receptionists were reduced [16], and patient histories did not need manual documentation [93]. Written queries were usually more detailed than when communicated verbally and were received by health care staff asynchronously, thus providing opportunities for more objective examination and more effective triage. Consequently, patient queries could more often be directed to other services or dealt with by other staff members rather than always by physicians [3]. Combined with their remote nature, OCs also gave staff more autonomy over how their work was organized, thus providing efficiency gains such as working from home and control over how to contact a patient rather than defaulting to an in-person consultation [13]. When telephone or in-person consultations were necessary, they were more focused and, therefore, quicker as the staff member could read the patient query before contact [17]. OCs reduced the workload for patients by avoiding the need to telephone their primary care provider to make an appointment, which often entailed long queues [18], and avoiding in-person consultations when possible, which typically involved travel, waiting rooms, and organizing time off work and childcare [15].

In contrast, 46% (29/63) of the studies suggested that OCs increased the workload for staff and patients. Staff described conducting OCs on top of their usual tasks [13] and dealing with them outside normal working hours [19]. They believed that, because OCs increased access to primary care, patients sought help more readily than they would have previously [17], thus creating “supply-induced demand” [108]. Processing OCs also created new administrative work such as filing them to EHRs and deciding whether they required input from a clinician [86]. Workload could also increase for patients if they perceived that entering their query into the OC system was more difficult than explaining it verbally [20].

OCs decreased costs for providers in 32% (20/63) of the studies largely by reducing in-person visits, which have associated expenditures related to staffing and utilities [21]. Patients reported that, owing to their convenience, having access to OCs stopped them from visiting other costly unscheduled care providers [92]. OCs decreased costs for patients in 6% (4/63) of the studies by avoiding in-person visits, which may entail expenses related to travel, unpaid work leave, and childcare [57].

In contrast, OCs increased costs for providers in 14% (9/63) of the studies owing to associated technology costs [63], time required for clinicians to triage patient queries [22], and insufficient reduction of in-person visits or telephone consultations [87].

Equitable

In all, 65% (41/63) of the studies suggested that OCs decreased equitable access to care services, as their use varied according to patient characteristics [63]. Conversely, 30% (19/63) of the studies suggested that OCs increased equitable access as they helped particular groups of patients who had previously struggled communicate with their primary care providers [14]. These characteristics are discussed in more detail in the Adopters section.

Patient-Centeredness

Although 21% (13/63) of the studies uncovered some patient dissatisfaction with OCs [90], 49% (31/63) found that most patients were at least as satisfied or more satisfied with OCs than with traditional in-person appointments [2]. Patients liked OCs for the aforementioned reasons: they improved access (timeliness), reduced their workload and costs (efficiency), and helped particular groups of patients communicate with their care providers (equitable).

Objective 2: How the Impacts of OCs on Primary Care Quality Are Influenced by System Design and Implementation

Condition Complexity

In all, 17% (11/63) of the studies suggested that OCs decreased staff workload when used for simple queries that were straightforward to resolve as they were more amenable to completion without needing to contact the patient directly via telephone or in person [5]. Simple queries included those related to administrative tasks, new and recurrent minor acute illnesses, prescriptions, tests, requests for advice, follow-up, and some chronic condition reviews [56]. These queries also decreased health costs as they saved clinicians time, for example, when administrative staff were able to relay messages and there was no direct contact between physician and patient [23]. In all, 11% (7/63) of the studies suggested that OCs increased staff workload and costs when used for complex queries such as those with multiple ill-defined symptoms [17]. These queries generally required verbal dialogue with and physical examination of the patient and were usually converted to telephone or in-person consultations to assess the patient further [23]. Staff felt that this duplicated the number of contacts with the patient for the same query.

Technology

In all, 21% (13/63) of the studies showed that, when patients had to use MCQs to input their OC query, it increased both patient and staff workload. Filling out long lists of questions shifted work from the clinician to the patient [20], and staff found them burdensome to read [86]. MCQs limited the amount of detail patients could enter, so staff could not always fully

understand their request. This increased workload as they often had to contact the patient to obtain further information [23]. MCQs also asked questions about seemingly "irrelevant" symptoms, which staff were responsible for assessing and following up, diverting attention from the patient's primary concern [10]. Owing to the restrictive nature of MCQs, patients regularly adapted their responses to obtain the outcome they wanted even when it was not the most appropriate use of resources. For example, reporting their symptoms differently to obtain an in-person consultation when self-care may have been more suitable ("gaming") [17].

In all, 14% (9/63) of the studies suggested that MCQs could also decrease patient satisfaction. Reasons included the amount of work required to complete them [14], their inflexibility in obtaining the answers patients wanted from their primary care provider [9], and that they could be confusing to navigate [25].

In contrast, 11% (7/63) of the studies suggested that, when patients could primarily report their queries using unstructured free text, it decreased staff workload and increased patient safety. This was because patients were more able to fully describe their query in sufficient detail using their own words, and clinicians did not have to request further information as often [95].

In 10% (6/63) of the studies, two-way written communication within the OC decreased the workload for both staff and patients. The ability to reply to patients in writing meant queries could be answered and follow-up questions could be asked at times convenient to both staff and patients, avoiding lengthy telephone and in-person consultations when appropriate [55]. It was also easier to communicate complex information, for example, by sending educational materials or using preset message templates [95].

In all, 21% (13/63) of the studies highlighted that a lack of integration between the OC system and other core software used by providers increased staff workload. Nonintegration meant that the staff had to go through multiple steps to perform a task, such as when filing an OC to a patient's EHR [21].

Adopters

Patients using OCs were more likely to be female (27/63, 43%) [70], younger (27/63, 43%) [91], and native speakers of the official language of the country they lived in (7/63, 11%) [25] and have a higher socioeconomic status (11/63, 17%) [57] than those not using OCs, thus decreasing equity. In contrast, both staff and patients felt that OCs increased access for particular groups of patients who struggled with traditional consultation methods, thus increasing equity and satisfaction with care. This included patients with mental health conditions who became anxious when speaking to health professionals on the telephone or in person (8/63, 13%) [20]; patients with verbal communication difficulties such as hearing loss who found it easier to communicate in writing (7/63, 11%) [90]; and patients with barriers to attending in-person appointments because of physical disabilities, geography, work commitments, or care responsibilities (8/63, 13%) [23]. In all, 13% (8/63) of the studies suggested that when staff and patients viewed traditional in-person methods as the gold standard, it could lead to resistance in adopting OCs [19].

Organization

In all, 8% (5/63) of the studies found that, when OCs were minimally advertised to patients, it understandably led to low rates of adoption [24]. In all, 11% (7/63) of the studies also showed that responding to a patient's initial OC query quickly led to high patient satisfaction, as it provided an advantage over traditional methods of primary care contact [6]; by definition, this also increased primary care access.

In all, 21% (13/63) of the studies found that the staff workload increased when providers did not integrate OCs into their normal daily workflows. For example, not scheduling time for clinicians to deal with OCs meant that they were done in addition to their normal tasks [93], and not diverting all incoming patient demand via the OC meant that different communication routes were often used for the same issue, thereby duplicating work [5]. In all, 13% (8/63) of the studies suggested that provider workload decreased if sufficient resources were allocated to implementing OCs. This included their initial setup—for example, training to enable staff to more effectively handle OCs [15]—and their ongoing processing—for example, dedicated facilities such as quiet rooms to help staff respond to OCs without distraction [55].

In all, 8% (5/63) of the studies showed that a lack of continuity of care between patients and their known physician negatively affected patient satisfaction. This occurred when any physician could reply to an OC query and patients were not able to specify a physician to whom to address their query [64].

Wider System

In all, 10% (6/63) of the studies showed that government policies mandating OC use increased their adoption. Example policies aimed to increase digital modes of contact with primary care in general [87] and minimize in-person contact during the COVID-19 pandemic [63]. In all, 8% (5/63) of the studies demonstrated that a lack of long-term external financial support for OCs limited their sustainability as health care organizations could often not afford to pay their ongoing costs [23].

Discussion

Summary of Evidence

This review focused on how OCs affect primary care quality, as defined by Institute of Medicine domains, for patients, providers, and the wider system, as well as which factors, as specified through the NASSS framework, influence this quality. We synthesized qualitative and quantitative evidence from 63 studies conducted in 9 countries covering 31 OC systems described in detail, with wide-ranging functionalities including AI. In all, 41% (26/63) of the studies were published in 2020 onward, and 17% (11/63) were published after the COVID-19 pandemic. Our main findings were that OCs are safe and have positive impacts on care quality, including increased access to care and decreased patient costs. However, they can have conflicting impacts on provider costs, staff and patient workloads, patient satisfaction, and care equity. We found that

the impacts OCs have on care quality are determined by the complexity of the patient queries they are used for, the design of the OC technology itself, the characteristics of staff and patient users, the way OCs are implemented by health care providers, and wider health policies.

Comparison of Findings With Other Reviews

Consistent with previous reviews relevant to OCs, we found a limited demographic of patients using OCs, leading to potential inequitable care [45,46]. We also found that the studies often did not sufficiently explore patients' perspectives of OCs in depth [46]; only 14% (9/63) of the studies used interview-based methods with an average sample size of 24.5 (SD 10.14). This hampered efforts to understand how such inequities arose.

Contrary to previous reviews, we found that OC impacts on care quality are more complex and nuanced than previously reported [45-47]. For example, we identified mixed findings regarding their impact on workload, patient satisfaction, and equitable care. This contrasts with previous reviews, where OCs only increased [47] or had no impact [45] on workload, decreased patient safety [45,47], and increased inequity [45-47].

These new findings for OCs may be partly explained because 76% (48/63) of the included studies had not been covered by these previous reviews. Although there was some overlap of papers (7/57, 12% of papers [45]; 7/13, 54% of papers [46]; and 4/17, 24% of papers [47]), most did not meet our inclusion criteria as they were either nonempirical (4/57, 7% [45]; 4/13, 31% [46]; and 4/17, 24% [47]), published before 2010 (26/57, 46% [45] and 2/17, 12% [47]), not based on real-world primary care (16/57, 28% [45]; 1/13, 8% [46]; and 6/17, 35% [47]), or did not meet our functional definition of an OC (39/57, 68% [45]; 2/13, 15% [46]; and 6/17, 35% [47]; eg, symptom checkers with no link to a health professional [28]).

By focusing on design and implementation, we identified new ways in which OCs affect primary care quality. For example, we found that, by increasing access, OCs can increase staff workload by creating "supply-induced demand" [17,108] and that they can decrease workload by enabling more focused consultations [17]. Furthermore, as previous reviews often did not analyze the design or implementation of OCs [45-47], we identified influential factors that have not been previously described. For example, although some reviews identified increased workload when clinicians received insufficient patient information via an OC system [46], we found that this was particularly associated with MCQ-based OCs [23]. We identified that allowing patients to describe their queries using unstructured free text had the opposite effect [95] while also having a positive impact on patient safety [55]. Using unstructured free text means that patients can more fully describe their query in addition to allowing them to freely express their ideas, concerns, and expectations, as is common in patient-centered primary care consultations [109].

Strengths and Limitations

As evidenced by the range of examples in Multimedia Appendix 1, we adopted a fundamental functional definition of OCs rather than relying on the names given to them by the authors of the included studies. When combined with our comprehensive searches across multiple databases and inclusion of gray literature, we identified more empirical studies relevant to OCs than any previous evidence synthesis on the topic [45-47]. Combined with our focus on causal mechanisms, this helped us develop a new and theoretically informed understanding of OCs that has not been previously reported.

As in all systematic reviews, our synthesis is reliant on what the study authors reported. OC features were not always described in sufficient detail to understand how they affected care quality [62]. There was also a lack of patient perspective in the studies, particularly from OC nonusers [4]. We made our literature search strategy as inclusive as possible regarding the different terms used for OCs (Multimedia Appendix 1) but, owing to their wide-ranging nature, it is possible that some papers were missed. We updated our searches between November 2021 and February 2022 to capture more recently published studies but, owing to time constraints, only 1 author (SD) screened these newer papers. This enabled us to capture studies conducted in the context of COVID-19 (11/63, 17% of all included studies).

Implications for Practice and Research

Overview

Our findings show that the impacts of OCs on care quality are complex and can be influenced by the subtle ways in which OCs are designed and implemented. To maximize their benefit for patients and staff, we therefore provide recommendations for OC developers on how systems could be designed, health care organizations on how they can be implemented and used, and researchers on questions and areas for further investigation. They are discussed in the following sections under the high-level themes from objective 2 and summarized in Table 4.

Table 4. Implications for online consultation (OC) research and practice.

Theme	Implications		
	OC designers	Health care providers	Researchers
Condition complexity	• Help health care providers identify when patients have submitted a query that could be unsuitable for resolution via an OC; for example, a complex condition	• Currently, all complex queries should be routed through traditional consultation methods	• Can OCs be used for complex queries and, if so, how can they be best adapted to support their resolution? • What impact do OCs have on clinical outcomes?
Technology	• Primarily allow patients to describe their queries using unstructured free text rather than MCQs[a] • Allow two-way written messages to be sent between staff and patients • Guide and support patients to provide sufficient detail about their query • Integrate with existing core clinical software systems used by health care organizations • Support patients to self-care or signpost them to other services when appropriate • Match capacity to demand by limiting the volume of OC queries a primary care provider can receive • Support workflow (eg, determining whether OCs need clinical vs administrative input) • Assist in triaging patient queries • Highlight when patients may require an in-person appointment • Explore the potential of using AI[b] to automate the aforementioned functions	• Guide and support patients to provide sufficient detail about their query	• Is the additional demand via OCs supply-induced or a previously unmet (and now unmasked) need? • How can AI be effectively used in OCs? • Fully describe the OC systems studied in detail (eg, using the TIDieR[c] checklist [110])
Adopters	• Involve patients from a variety of backgrounds in designing OC systems to facilitate their adoption	• Involve patients from a variety of backgrounds in planning how OCs are implemented • Explain and promote the benefits of OCs to staff and patients during their implementation—including increased access for certain patient groups (eg, those with mental health conditions, verbal communication difficulties, and barriers to attending in-person appointments)	• What is the experience of patient users and low or nonusers of OCs from a range of backgrounds? • Why are patients with different characteristics more or less likely to use OCs? • How can patients from different backgrounds be supported to use OCs effectively? • Are there other specific patient groups likely to benefit from OCs and why? • In what circumstances are in-person consultation methods viewed as the gold standard and why? • How are OCs being used after the COVID-19 pandemic?
Organization	• Facilitate planning and booking OCs into clinicians' daily schedules	• Widely promote OCs to patients through various channels (eg, mail-out campaigns) • Provide sufficient staff training on OCs • Divert as much incoming patient demand as possible through OCs • Plan OCs into clinicians' daily schedules • Initially respond to patients through written message or phone call as soon as possible on the same day to acknowledge their query	• How can OCs most effectively be incorporated into daily workflows? • Are OCs suitable for middle-income countries?

Theme	Implications		
	OC designers	Health care providers	Researchers
Wider system	N/A[d]	• Use system-wide policies to increase OC uptake • Centralized funding is required to ensure sustainability	• What is the long-term experience of policies mandating OC use, particularly in light of the COVID-19 pandemic?

[a]MCQ: multiple-choice questionnaire.

[b]AI: artificial intelligence.

[c]TIDieR: Template for Intervention Description and Replication.

[d]N/A: not applicable.

Condition Complexity

It is unclear whether OCs are unsuitable for complex patient queries or whether workflows and procedures can be better organized and OC systems can be better designed to deal with them. Therefore, we recommend that (1) complex conditions are routed through traditional consultation methods (eg, in person and telephone) and (2) further research is conducted on how these types of conditions could be better handled via OCs to ensure that they benefit all patients.

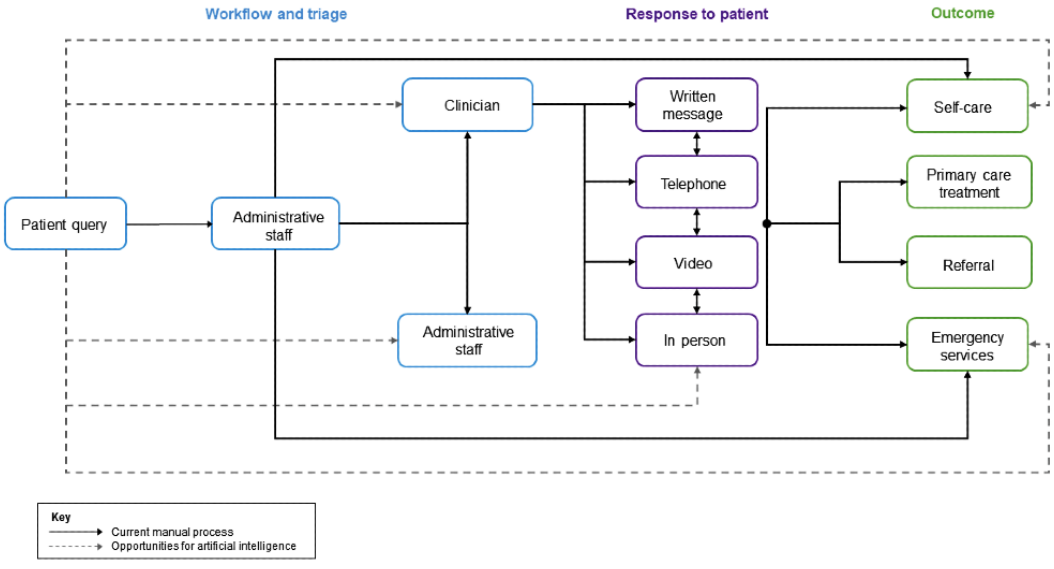
Technology

On the basis of existing evidence, we recommend that OC developers (1) allow patients to fully describe their queries using unstructured free text rather than MCQs, (2) support patients in providing sufficient detail in their queries for their primary care provider to respond quickly and safely, (3) allow for two-way written communication between staff and patients, and (4) integrate their solutions with existing core clinical software systems.

Technology design also plays a role in mitigating some of the undesirable outcomes we identified from using OCs, including increasing workload and costs. Increased workload is particularly important as it can lead to a mismatch between patient demand and health care resources, which can in turn threaten patient safety if providers are unable to deal with OCs in an appropriate time frame. A way this could happen is through increased demand—if there are too many OCs submitted by patients and not enough staff to deal with them [55]. Whether this additional demand is a supply-induced [108] or previously unmet (and now unmasked) need was unclear from the studies we included [15] and requires further research. Nevertheless, OC systems could help by (1) supporting patients to self-care or signposting them to other services when appropriate; (2) matching capacity to demand by limiting the number of OC queries that primary care providers can receive from patients; (3) supporting workflow, for example, by determining whether OCs require clinical input to relieve the workload of administrators [86]; (4) assisting in triaging patient queries to reduce the associated costs of solely relying on clinicians for triage [22]; and (5) highlighting when patients may require an in-person appointment to facilitate direct booking to avoid work duplication [23], which may relate to patient query complexity.

According to our definition [82], many of these functions may require AI to be most effective, which should be explored by OC designers (Figure 2). In all, 54% (13/24) of MCQ-based OC systems in our review used AI (Table 1) [54], although largely for other functions rather than the aforementioned ones. Furthermore, AI was usually not the focus of the studies, and we consequently found only low-confidence evidence regarding its use in OCs (Multimedia Appendix 8). Therefore, how AI could be used by OC systems in clinical practice requires further research.

Figure 2. Artificial intelligence opportunities.



The included papers did not always adequately describe the OC systems studied, limiting our ability to determine how their specific features affected care quality. Future research should describe OC systems in detail so that evaluation findings can be usefully compared, for example, by using the Template for Intervention Description and Replication checklist [110].

Adopters

We found inadequate exploration of participant (especially patient) experiences to confidently explain how and why the impacts on care equity arose during OC use. Study authors and health care staff often speculated reasons [18], but this was insufficient to formulate evidence-based hypotheses. Future research should explore the perspectives of patients using (and not using) OCs from a wide range of backgrounds using in-depth qualitative techniques such as interview-based methods. Patients from a variety of backgrounds should be involved in how OC systems are designed and help plan how they are implemented in practice.

Staff and patients resisted adopting OCs when they viewed traditional in-person consultation methods as the gold standard. Although this was understandable for complex queries [17], it was unclear whether other factors also influenced this view. Future research should address this evidence gap, particularly as COVID-19 has made remote consultations more commonplace [49]. In the meantime, this perception could be challenged by explaining the benefits of OCs found in our review to prospective users [111].

Organization

For patients and staff to experience the benefits of OCs, they must be widely promoted to patients as a route for them to contact their primary care provider. This can happen through various channels, such as mail-out campaigns (eg, via SMS text message) or by verbally mentioning OCs when in contact with patients (eg, when receptionists speak to patients on the telephone).

To minimize workload associated with OCs, we recommend that organizations (1) allocate sufficient resources to both setting up and processing them, including the provision of training on how to use OCs, and to staff and facilities (eg, computers and rooms) to deal with them; (2) divert as much incoming patient demand as possible through the system to avoid duplication and increase the proportion of patient contacts that benefit from OCs; and (3) incorporate OCs into daily work patterns by scheduling protected time for staff to deal with them to ensure that they do not become additional tasks to complete on top of their normal work.

Our findings show that providers can increase access and patient satisfaction by responding quickly to OCs, although the definitions of what this involved were unclear. We recommend providing an initial response to patients' OC queries as soon as possible on the same day—either through written message or telephone call. This does not mean that the entire query needs to be resolved at this point, only that initial contact has been made and the query has been acknowledged.

We included studies from 9 countries, all of which were high-income Western countries. Owing to their remote nature, OCs may play a role in middle-income countries where there are isolated communities and fewer health care staff per head of population. However, further research is required to understand how their technological and financial barriers could be overcome.

Wider System

Governmental policies to promote OCs are effective in increasing adoption, although centralized funding is needed to sustain their use. It is unclear what the long-term experience of such policies is from the papers we included, particularly in response to those relating to the COVID-19 pandemic.

Conclusions

This is the first theoretically informed synthesis of empirical research on OCs in primary care and uniquely includes studies conducted during the COVID-19 pandemic. It contributes new knowledge that OCs are safe and have positive impacts on care quality, including increased access to primary care and decreased patient costs. However, they are also complex and often produce conflicting impacts on provider costs, staff and patient workloads, patient satisfaction, and care equity. Some of these are unintended and conflict with the promotion of OCs by policy makers as a way to address already increasing workload and decreasing workforce capacity in primary care [31-36]. Unlike previous evidence syntheses on the topic, we have shown that negative impacts on care quality of OCs can be mitigated through appropriate system design (eg, free text formats and two-way written communication), incorporation of advanced technologies (eg, AI), and integration into technical infrastructure (eg, EHRs) and organizational workflows (eg, timely responses). Since the advent of COVID-19, OCs have become indispensable, although further engineering and implementation research is required to realize their full benefits.

Acknowledgments

This research was funded by Innovate UK (105178) and a Wellcome Trust Clinical Research Career Development Fellowship for BCB (209593/Z/17/Z). NP's time was partially funded by the National Institute for Health and Care Research (NIHR) Greater Manchester Patient Safety Translational Research Centre. 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. The NIHR had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data and the final responsibility to submit for publication.

Authors' Contributions

SD, TC, and BCB refined the research question; developed the search strings; conducted screening, critical appraisal, data extraction, and data analysis; and wrote the first draft of the manuscript. NP contributed to the conception and design of the review. All authors contributed to the final analysis and approved the final submitted version of the manuscript.

Conflicts of Interest

BCB is clinical lead for a commercially available online consultation system.

Multimedia Appendix 1
Terms used by the included studies for online consultations.
[DOCX File , 25 KB - jmir_v24i10e37436_app1.docx]

Multimedia Appendix 2
Search terms.
[DOC File , 41 KB - jmir_v24i10e37436_app2.doc]

Multimedia Appendix 3
Data extraction form.
[DOC File , 154 KB - jmir_v24i10e37436_app3.doc]

Multimedia Appendix 4
Descriptive summary of the included studies.
[DOC File , 185 KB - jmir_v24i10e37436_app4.doc]

Multimedia Appendix 5
Description of the online consultation systems studied.
[DOC File , 128 KB - jmir_v24i10e37436_app5.doc]

Multimedia Appendix 6
Quality appraisal of the included studies using the Mixed Methods Appraisal Tool.
[DOC File , 331 KB - jmir_v24i10e37436_app6.doc]

Multimedia Appendix 7
Low-confidence findings for objective 1.
[DOC File , 41 KB - jmir_v24i10e37436_app7.doc]

Multimedia Appendix 8
Low-confidence findings for objective 2.
[DOCX File , 57 KB - jmir_v24i10e37436_app8.docx]

Multimedia Appendix 9
Outcomes of online consultations in primary care (with exemplar data).
[DOCX File , 25 KB - jmir_v24i10e37436_app9.docx]

Multimedia Appendix 10
How outcomes of online consultations in primary care are influenced by system design and implementation (with exemplar data).
[DOCX File , 49 KB - jmir_v24i10e37436_app10.docx]

References

1. Bakhai M, Atherton H. How to conduct written online consultations with patients in primary care. BMJ 2021 Feb 24;372:n264. [doi: 10.1136/bmj.n264] [Medline: 33627324]
2. Johansson A, Larsson M, Ivarsson B. Patients' experiences with a digital primary health care concept using written dialogues: a pilot study. J Prim Care Community Health 2020;11:2150132720910564 [FREE Full text] [doi: 10.1177/2150132720910564] [Medline: 32114868]

3. Johansson A, Larsson M, Ivarsson B. General practitioners' experiences of digital written patient dialogues: a pilot study using a mixed method. J Prim Care Community Health 2020;11:2150132720909656 [FREE Full text] [doi: 10.1177/2150132720909656] [Medline: 32133905]
4. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Feb 25;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
5. Eldh AC, Sverker A, Bendtsen P, Nilsson E. Health care professionals' experience of a digital tool for patient exchange, anamnesis, and triage in primary care: qualitative study. JMIR Hum Factors 2020 Dec 14;7(4):e21698 [FREE Full text] [doi: 10.2196/21698] [Medline: 33315014]
6. Leung K, Qureshi S. Managing high frequency users of an electronic consultation system in primary care: a quality improvement project. BMJ Open Qual 2021 Jun;10(2):e001310 [FREE Full text] [doi: 10.1136/bmjoq-2020-001310] [Medline: 34112657]
7. Landgren S, Cajander Å. Non-use of digital health consultations among Swedish elderly living in the countryside. Front Public Health 2021 Sep 10;9:588583 [FREE Full text] [doi: 10.3389/fpubh.2021.588583] [Medline: 34568247]
8. Nijhof D, Ingram A, Ochieng R, Roberts EJ, Poulton B, Ochieng B. Examining GP online consultation in a primary care setting in east midlands, UK. BMC Health Serv Res 2021 Sep 30;21(1):1030 [FREE Full text] [doi: 10.1186/s12913-021-07039-2] [Medline: 34592980]
9. Nilsson E, Sverker A, Bendtsen P, Eldh AC. A human, organization, and technology perspective on patients' experiences of a chat-based and automated medical history-taking service in primary health care: interview study among primary care patients. J Med Internet Res 2021 Oct 18;23(10):e29868 [FREE Full text] [doi: 10.2196/29868] [Medline: 34661544]
10. Cajander Å, Larusdottir M, Hedström G. The effects of automation of a patient-centric service in primary care on the work engagement and exhaustion of nurses. Qual User Exp 2020 Sep 19;5(1):9. [doi: 10.1007/s41233-020-00038-x]
11. Zanaboni P, Fagerlund AJ. Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. BMJ Open 2020 Jun 17;10(6):e034773 [FREE Full text] [doi: 10.1136/bmjopen-2019-034773] [Medline: 32554721]
12. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Evaluation of the use of an ask-the-expert e-consultation service for support on health-related requests. In: Proceedings of the 2nd International Conference on eHealth, Telemedicine, and Social Medicine. 2010 Presented at: eTELEMED '10; February 10-16, 2010; Saint Maarten, The Netherlands p. 72-76. [doi: 10.1109/etelemed.2010.33]
13. Wästfelt E, Peber E. Impact of digi-physical healthcare. Lund University. 2020. URL: https://lup.lub.lu.se/student-papers/search/publication/9015211 [accessed 2020-09-01]
14. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]
15. Evaluation of Babylon GP at Hand: final evaluation report. Ipsos MORI and York Health Economics Consortium. 2019 May. URL: http://allcatsrgrey.org.uk/wp/download/informatics/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf [accessed 2020-09-01]
16. Fagerlund AJ, Holm IM, Zanaboni P. General practitioners' perceptions towards the use of digital health services for citizens in primary care: a qualitative interview study. BMJ Open 2019 May 05;9(5):e028251 [FREE Full text] [doi: 10.1136/bmjopen-2018-028251] [Medline: 31061056]
17. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8 [FREE Full text] [doi: 10.3399/bjgp17X693509] [Medline: 29109115]
18. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. Health Serv Deliv Res 2018 Jun;6(20):1-200. [doi: 10.3310/hsdr06200] [Medline: 29889485]
19. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018 Apr;68(669):e293-e300 [FREE Full text] [doi: 10.3399/bjgp18X694853] [Medline: 29378697]
20. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):e018688 [FREE Full text] [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]
21. Matheson C. Implementation of WebGP and E-consultations in Wessex GP Practices: Interim Update Report. University of Southampton. Southampton, UK: Centre of Implementation Science; 2016 Jun 15. URL: https://tinyurl.com/2h9s9pb5 [accessed 2020-09-01]
22. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]

23. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Mar 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
24. Greater Manchester digital primary care insight full report: Key findings from engagement June-July 2020. Health Innovation Manchester. 2020. URL: https://healthinnovationmanchester.com/wp-content/uploads/2020/12/PCIE-DPC-Full-report-FINAL-20.10.20-1.pdf [accessed 2021-10-01]
25. GM Digital First Primary Care: Patient and public insights: Workshop results. Health Innovation Manchester. 2020. URL: https://healthinnovationmanchester.com/our-work/gm-digital-first-primary-care/ [accessed 2021-09-01]
26. Bertelsen P, Petersen LS. Danish citizens and general practitioners' use of ICT for their mutual communication. In: Proceedings of the 15th World Congress on Health and Biomedical Informatics. 2015 Presented at: MEDINFO '15; August 19-23, 2015; São Paulo, Brazil p. 367-369. [doi: 10.3233/978-1-61499-564-7-376]
27. Ekman B, Thulesius H, Wilkens J, Lindgren A, Cronberg O, Arvidsson E. Utilization of digital primary care in Sweden: descriptive analysis of claims data on demographics, socioeconomics, and diagnoses. Int J Med Inform 2019 Jul;127:134-140. [doi: 10.1016/j.ijmedinf.2019.04.016] [Medline: 31128825]
28. 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]
29. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
30. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
31. The NHS Long Term Plan. National Health Service. 2019 Jan 7. URL: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2021-03-01]
32. Telehealth and technology-based health services in primary care. The Royal New Zealand College Of General Practitioners. 2017 Nov. URL: https://www.rnzcgp.org.nz/gpdocs/New-website/Advocacy/Position-Statements/Telehealth-and-technology-based-health-services-in-primary-care-updated-....pdf [accessed 2021-03-01]
33. Gill M. A National Telehealth Strategy For Australia – For Discussion. Australian National Consultative Committee on Electronic Health. 2011 Nov. URL: https://globalaccesspartners.org/ancceh_telehealth_paper.pdf [accessed 2022-03-01]
34. Daniel H, Sulmasy LS, Health and Public Policy Committee of the American College of Physicians. Policy recommendations to guide the use of telemedicine in primary care settings: an American College of Physicians position paper. Ann Intern Med 2015 Nov 17;163(10):787-789 [FREE Full text] [doi: 10.7326/M15-0498] [Medline: 26344925]
35. Hobbs FD, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet 2016 Jun 04;387(10035):2323-2330 [FREE Full text] [doi: 10.1016/S0140-6736(16)00620-6] [Medline: 27059888]
36. Marchand C, Peckham S. Addressing the crisis of GP recruitment and retention: a systematic review. Br J Gen Pract 2017 Apr;67(657):e227-e237 [FREE Full text] [doi: 10.3399/bjgp17X689929] [Medline: 28289014]
37. Beaney P, Odulaja A, Hadley A, Prince C, Obe RC. GP Online: turning expectations into reality with the new NHS app. Br J Gen Pract 2019 Apr;69(681):172-173 [FREE Full text] [doi: 10.3399/bjgp19X701333] [Medline: 30745356]
38. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021 Jun;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
39. Joy M, McGagh D, Jones N, Liyanage H, Sherlock J, Parimalanathan V, et al. Reorganisation of primary care for older adults during COVID-19: a cross-sectional database study in the UK. Br J Gen Pract 2020 Aug;70(697):e540-e547 [FREE Full text] [doi: 10.3399/bjgp20X710933] [Medline: 32661009]
40. Han SM, Greenfield G, Majeed A, Hayhoe B. Impact of remote consultations on antibiotic prescribing in primary health care: systematic review. J Med Internet Res 2020 Nov 09;22(11):e23482 [FREE Full text] [doi: 10.2196/23482] [Medline: 33031045]
41. 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]
42. 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]
43. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
44. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

45. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
46. Baines R, Tredinnick-Rowe J, Jones R, Chatterjee A. Barriers and enablers in implementing electronic consultations in primary care: scoping review. J Med Internet Res 2020 Nov 12;22(11):e19375 [FREE Full text] [doi: 10.2196/19375] [Medline: 33035177]
47. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
48. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
49. Neves AL, Li E, Gupta PP, Fontana G, Darzi A. Virtual primary care in high-income countries during the COVID-19 pandemic: policy responses and lessons for the future. Eur J Gen Pract 2021 Dec;27(1):241-247 [FREE Full text] [doi: 10.1080/13814788.2021.1965120] [Medline: 34431426]
50. Bakhai M, Croney L, Waller O, Henshall N, Felstead C. Using online consultations in primary care: implementation toolkit. National Health Service England. 2019 Sep 26. URL: https://www.england.nhs.uk/publication/using-online-consultations-in-primary-care-implementation-toolkit/ [accessed 2020-01-14]
51. Marshall M, Shah R, Stokes-Lampard H. Online consulting in general practice: making the move from disruptive innovation to mainstream service. BMJ 2018 Mar 26;360:k1195. [doi: 10.1136/bmj.k1195] [Medline: 29581174]
52. Stamenova V, Agarwal P, Kelley L, Fujioka J, Nguyen M, Phung M, et al. Uptake and patient and provider communication modality preferences of virtual visits in primary care: a retrospective cohort study in Canada. BMJ Open 2020 Jul 06;10(7):e037064 [FREE Full text] [doi: 10.1136/bmjopen-2020-037064] [Medline: 32636284]
53. Fernández OS, Seguí FL, Vidal-Alaball J, Bonet Simo JM, Vian OH, Cabo PR, et al. Primary care doctor characteristics that determine the use of teleconsultations in the Catalan public health system: retrospective descriptive cross-sectional study. JMIR Med Inform 2020 Jan 31;8(1):e16484 [FREE Full text] [doi: 10.2196/16484] [Medline: 32012061]
54. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
55. Entezarjou A, Bolmsjö BB, Calling S, Midlöv P, Milos Nymberg V. Experiences of digital communication with automated patient interviews and asynchronous chat in Swedish primary care: a qualitative study. BMJ Open 2020 Jul 23;10(7):e036585 [FREE Full text] [doi: 10.1136/bmjopen-2019-036585] [Medline: 32709650]
56. López Seguí F, Walsh S, Solans O, Adroher Mas C, Ferraro G, García-Altés A, et al. Teleconsultation between patients and health care professionals in the Catalan primary care service: message annotation analysis in a retrospective cross-sectional study. J Med Internet Res 2020 Sep 17;22(9):e19149 [FREE Full text] [doi: 10.2196/19149] [Medline: 32687477]
57. Kelley LT, Phung M, Stamenova V, Fujioka J, Agarwal P, Onabajo N, et al. Exploring how virtual primary care visits affect patient burden of treatment. Int J Med Inform 2020 Sep;141:104228. [doi: 10.1016/j.ijmedinf.2020.104228] [Medline: 32683311]
58. López Seguí F, Vidal-Alaball J, Sagarra Castro M, García-Altés A, García Cuyàs F. General practitioners' perceptions of whether teleconsultations reduce the number of face-to-face visits in the Catalan public primary care system: retrospective cross-sectional study. J Med Internet Res 2020 Mar 16;22(3):e14478 [FREE Full text] [doi: 10.2196/14478] [Medline: 32175914]
59. Murray MA, Penza KS, Myers JF, Furst JW, Pecina JL. Comparison of eVisit management of urinary symptoms and urinary tract infections with standard care. Telemed J E Health 2020 May;26(5):639-644. [doi: 10.1089/tmj.2019.0044] [Medline: 31313978]
60. Penza KS, Murray MA, Myers JF, Furst JW, Pecina JL. Management of acute sinusitis via e-visit. Telemed J E Health 2021 May;27(5):532-536. [doi: 10.1089/tmj.2020.0047] [Medline: 32522103]
61. Entezarjou A, Bonamy AK, Benjaminsson S, Herman P, Midlöv P. Human- versus machine learning-based triage using digitalized patient histories in primary care: comparative study. JMIR Med Inform 2020 Sep 03;8(9):e18930 [FREE Full text] [doi: 10.2196/18930] [Medline: 32880578]
62. Tarn DM, Hintz C, Mendez-Hernandez E, Sawlani SP, Bholat MA. Using virtual visits to care for primary care patients with COVID-19 symptoms. J Am Board Fam Med 2021 Feb;34(Suppl):S147-S151 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200241] [Medline: 33622830]
63. Wilson G, Currie O, Bidwell S, Saeed B, Dowell A, Halim AA, et al. Empty waiting rooms: the New Zealand general practice experience with telehealth during the COVID-19 pandemic. N Z Med J 2021 Jul 09;134(1538):89-101. [Medline: 34239148]
64. Turner A, Morris R, Rakhra D, Stevenson F, McDonagh L, Hamilton F, et al. Unintended consequences of online consultations: a qualitative study in UK primary care. Br J Gen Pract 2022 Jan 27;72(715):e128-e137 [FREE Full text] [doi: 10.3399/BJGP.2021.0426] [Medline: 34903520]
65. Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA, USA: Sage Publications; 1997.

66. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health 2019 Jan 25;4(Suppl 1):e000893 [FREE Full text] [doi: 10.1136/bmjgh-2018-000893] [Medline: 30775016]
67. Systematic Reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination. York, UK: University of York; 2009 Jan. URL: https://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm [accessed 2021-01-03]
68. 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]
69. Darley S, Coulson T, Peek N, Vanderveer S, Wong D, Brown B. Online patient triage and consultation in primary care: a systematic review – CRD42020191802. PROSPERO. 2020. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020191802 [accessed 2022-09-06]
70. Penza KS, Murray MA, Pecina JL, Myers JF, Furst JW. Electronic visits for minor acute illnesses: analysis of patient demographics, prescription rates, and follow-up care within an asynchronous text-based online visit. Telemed J E Health 2018 Mar;24(3):210-215. [doi: 10.1089/tmj.2017.0091] [Medline: 28731388]
71. Albert SM, Shevchik GJ, Paone S, Martich GD. Internet-based medical visit and diagnosis for common medical problems: experience of first user cohort. Telemed J E Health 2011 May;17(4):304-308 [FREE Full text] [doi: 10.1089/tmj.2010.0156] [Medline: 21457013]
72. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
73. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. Characteristics of patients who seek care via eVisits instead of office visits. Telemed J E Health 2013 Jul;19(7):515-519 [FREE Full text] [doi: 10.1089/tmj.2012.0221] [Medline: 23682589]
74. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
75. Abimbola S, Patel B, Peiris D, Patel A, Harris M, Usherwood T, et al. The NASSS framework for ex post theorisation of technology-supported change in healthcare: worked example of the TORPEDO programme. BMC Med 2019 Dec 30;17(1):233 [FREE Full text] [doi: 10.1186/s12916-019-1463-x] [Medline: 31888718]
76. 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]
77. Hong QN. Reporting the results of the MMAT (version 2018). Mixed Methods Appraisal Tool. 2020 Dec 9. URL: https://tinyurl.com/y9ebmcp8 [accessed 2021-03-01]
78. NVivo: qualitative research data analysis software. QSR International. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/try-nvivo [accessed 2020-08-01]
79. Stern C, Lizarondo L, Carrier J, Godfrey C, Rieger K, Salmond S, et al. Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evid Synth 2020 Oct;18(10):2108-2118. [doi: 10.11124/JBISRIR-D-19-00169] [Medline: 32813460]
80. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
81. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC, USA: National Academies Press (US); 2001.
82. Chen L, Wong G. Transcriptome informatics. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of Bioinformatics and Computational Biology. Oxford, UK: Academic Press; 2019:324-340.
83. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data. London, UK: Routledge; 1994:173-194.
84. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 2015 Oct 27;12(10):e1001895 [FREE Full text] [doi: 10.1371/journal.pmed.1001895] [Medline: 26506244]
85. Online consultations research: Summary research findings. National Health Service England and National Health Service Improvement. 2019 Feb. URL: https://www.england.nhs.uk/wp-content/uploads/2019/09/online-consultations-reserach-summary-of-findings.pdf [accessed 2020-09-01]
86. Casey M, Shaw S, Swinglehurst D. Experiences with online consultation systems in primary care: case study of one early adopter site. Br J Gen Pract 2017 Nov;67(664):e736-e743 [FREE Full text] [doi: 10.3399/bjgp17X693137] [Medline: 28993306]
87. Andersen KN, Nielsen JA, Kim S. Use, cost, and digital divide in online public health care: lessons from Denmark. Transform Gov People Process Policy 2019 May 20;13(2):197-211. [doi: 10.1108/tg-06-2018-0041]

88. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
89. Player M, O'Bryan E, Sederstrom E, Pinckney J, Diaz V. Electronic visits for common acute conditions: evaluation of a recently established program. Health Aff (Millwood) 2018 Dec;37(12):2024-2030. [doi: 10.1377/hlthaff.2018.05122] [Medline: 30633677]
90. Eccles A, Hopper M, Turk A, Atherton H. Patient use of an online triage platform: a mixed-methods retrospective exploration in UK primary care. Br J Gen Pract 2019 May;69(682):e336-e344 [FREE Full text] [doi: 10.3399/bjgp19X702197] [Medline: 30910874]
91. Jung C, Padman R. Virtualized healthcare delivery: understanding users and their usage patterns of online medical consultations. Int J Med Inform 2014 Dec;83(12):901-914. [doi: 10.1016/j.ijmedinf.2014.08.004] [Medline: 25193501]
92. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
93. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]
94. Padman R, Shevchik G, Paone S, Dolezal C, Cervenak J. eVisit: a pilot study of a new kind of healthcare delivery. Stud Health Technol Inform 2010;160(Pt 1):262-266. [Medline: 20841690]
95. Adamson SC, Bachman JW. Pilot study of providing online care in a primary care setting. Mayo Clin Proc 2010 Aug;85(8):704-710 [FREE Full text] [doi: 10.4065/mcp.2010.0145] [Medline: 20516427]
96. Lawless M, Wright E, Davidson J. A collaborative approach to improving patient access in general practice: impact of three different pilot schemes in 12 general practices in Greenwich. London J Prim Care (Abingdon) 2016 Jun 10;8(4):56-65 [FREE Full text] [doi: 10.1080/17571472.2016.1173946] [Medline: 28250835]
97. Johnson KM, Dumkow LE, Burns KW, Yee MA, Egwuatu NE. Comparison of diagnosis and prescribing practices between virtual visits and office visits for adults diagnosed with sinusitis within a primary care network. Open Forum Infect Dis 2019 Oct 5;6(9):ofz393 [FREE Full text] [doi: 10.1093/ofid/ofz393] [Medline: 31660415]
98. Bavafa H, Hitt LM, Terwiesch C. The impact of e-Visits on visit frequencies and patient health: evidence from primary care. Manage Sci 2018 Dec;64(12):5461-5480 [FREE Full text] [doi: 10.1287/mnsc.2017.2900] [Medline: 33033417]
99. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. JAMA Intern Med 2013 Jan 14;173(1):72-74 [FREE Full text] [doi: 10.1001/2013.jamainternmed.305] [Medline: 23403816]
100. Rohrer JE, Angstman KB, Adamson SC, Bernard ME, Bachman JW, Morgan ME. Impact of online primary care visits on standard costs: a pilot study. Popul Health Manag 2010 Apr;13(2):59-63. [doi: 10.1089/pop.2009.0018] [Medline: 20415617]
101. Jung C, Padman R, Shevchik G, Paone S. Who are portal users vs. early e-Visit adopters? A preliminary analysis. AMIA Annu Symp Proc 2011;2011:1070-1079 [FREE Full text] [Medline: 22195168]
102. Prime Minister's Challenge Fund: Improving Access to General Practice. First Evaluation Report: October 2015. National Health Service England. 2015 Oct. URL: https://www.england.nhs.uk/wp-content/uploads/2015/10/pmcf-wv-one-eval-report.pdf [accessed 2020-09-01]
103. Hertzog R, Johnson J, Smith J, McStay FW, da Graca B, Haneke T, et al. Diagnostic accuracy in primary care e-visits: evaluation of a large integrated health care delivery system's experience. Mayo Clin Proc 2019 Jun;94(6):976-984. [doi: 10.1016/j.mayocp.2019.02.011] [Medline: 31171135]
104. Peabody MR, Dai M, Turner K, Peterson LE, Mainous 3rd AG. Prevalence and factors associated with family physicians providing e-Visits. J Am Board Fam Med 2019;32(6):868-875 [FREE Full text] [doi: 10.3122/jabfm.2019.06.190081] [Medline: 31704755]
105. Leng S, MacDougall M, McKinstry B. The acceptability to patients of video-consulting in general practice: semi-structured interviews in three diverse general practices. J Innov Health Inform 2016 Jul 15;23(2):141 [FREE Full text] [doi: 10.14236/jhi.v23i2.141] [Medline: 27869579]
106. Ruas SS, Assunção A. Teleconsultations by primary care physicians of Belo Horizonte: challenges in the diffusion of innovation. Telemed J E Health 2013 May;19(5):409-414. [doi: 10.1089/tmj.2012.0095] [Medline: 23600411]
107. Jury SC, Kornberg AJ. Integrating telehealth in to 'business as usual': Is it really possible? J Telemed Telecare 2016 Dec;22(8):499-503. [doi: 10.1177/1357633X16675802] [Medline: 27799455]
108. Rosen R. Meeting need or fueling demand? Improved access to primary care and supply-induced demand. Nuffield Trust and National Health Service England. 2014 Jun. URL: https://www.nuffieldtrust.org.uk/files/2017-01/meeting-need-or-fuelling-demand-web-final.pdf [accessed 2020-09-01]
109. Larsen JH, Neighbour R. Five cards: a simple guide to beginning the consultation. Br J Gen Pract 2014 Mar;64(620):150-151 [FREE Full text] [doi: 10.3399/bjgp14X677662] [Medline: 24567647]
110. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]

111. Rogers EM. Diffusion of Innovations. New York, NY, USA: Free Press; 1962.

Abbreviations

AI: artificial intelligence
EHR: electronic health record
MCQ: multiple-choice questionnaire
MMAT: Mixed Methods Appraisal Tool
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability framework
OC: online consultation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by C Basch; submitted 22.02.22; peer-reviewed by A Entezarjou, T Ong; comments to author 18.04.22; revised version received 29.05.22; accepted 31.05.22; published 24.10.22.

Please cite as:
Darley S, Coulson T, Peek N, Moschogianis S, van der Veer SN, Wong DC, Brown BC
Understanding How the Design and Implementation of Online Consultations Affect Primary Care Quality: Systematic Review of Evidence With Recommendations for Designers, Providers, and Researchers
J Med Internet Res 2022;24(10):e37436
URL: https://www.jmir.org/2022/10/e37436
doi:10.2196/37436
PMID:36279172

Review

Visualization Techniques of Time-Oriented Data for the Comparison of Single Patients With Multiple Patients or Cohorts: Scoping Review

Jan Scheer[1], MSc; Alisa Volkert[2], MSc; Nicolas Brich[3], MSc; Lina Weinert[4], MSc; Nandhini Santhanam[5], MSc; Michael Krone[3], PhD; Thomas Ganslandt[6], PhD; Martin Boeker[7], PhD; Till Nagel[1], PhD

[1]Human Data Interaction Lab, Mannheim University of Applied Sciences, Mannheim, Germany

[2]Center for Innovative Care, University Hospital Tübingen, Tübingen, Germany

[3]Big Data Visual Analytics in Life Sciences, Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany

[4]Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

[5]Abteilung für Biomedizinische Informatik, Zentrum für Präventivmedizin und Digitale Gesundheit Baden-Württemberg, Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Germany

[6]Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[7]University Hospital Rechts der Isar, School of Medicine, Technical University of Munich, München, Germany

Corresponding Author:
Till Nagel, PhD
Human Data Interaction Lab
Mannheim University of Applied Sciences
Paul-Wittsack-Straße 10
Mannheim, 68163
Germany
Phone: 49 621 292 6886
Email: t.nagel@hs-mannheim.de

Abstract

Background: Visual analysis and data delivery in the form of visualizations are of great importance in health care, as such forms of presentation can reduce errors and improve care and can also help provide new insights into long-term disease progression. Information visualization and visual analytics also address the complexity of long-term, time-oriented patient data by reducing inherent complexity and facilitating a focus on underlying and hidden patterns.

Objective: This review aims to provide an overview of visualization techniques for time-oriented data in health care, supporting the comparison of patients. We systematically collected literature and report on the visualization techniques supporting the comparison of time-based data sets of single patients with those of multiple patients or their cohorts and summarized the use of these techniques.

Methods: This scoping review used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. After all collected articles were screened by 16 reviewers according to the criteria, 6 reviewers extracted the set of variables under investigation. The characteristics of these variables were based on existing taxonomies or identified through open coding.

Results: Of the 249 screened articles, we identified 22 (8.8%) that fit all criteria and reviewed them in depth. We collected and synthesized findings from these articles for medical aspects such as medical context, medical objective, and medical data type, as well as for the core investigated aspects of visualization techniques, interaction techniques, and supported tasks. The extracted articles were published between 2003 and 2019 and were mostly situated in clinical research. These systems used a wide range of visualization techniques, most frequently showing changes over time. Timelines and temporal line charts occurred 8 times each, followed by histograms with 7 occurrences and scatterplots with 5 occurrences. We report on the findings quantitatively through visual summarization, as well as qualitatively.

Conclusions: The articles under review in general mitigated complexity through visualization and supported diverse medical objectives. We identified 3 distinct patient entities: single patients, multiple patients, and cohorts. Cohorts were typically visualized in condensed form, either through prior data aggregation or through visual summarization, whereas visualization of individual

patients often contained finer details. All the systems provided mechanisms for viewing and comparing patient data. However, explicitly comparing a single patient with multiple patients or a cohort was supported only by a few systems. These systems mainly use basic visualization techniques, with some using novel visualizations tailored to a specific task. Overall, we found the visual comparison of measurements between single and multiple patients or cohorts to be underdeveloped, and we argue for further research in a systematic review, as well as the usefulness of a design space.

(J Med Internet Res 2022;24(10):e38041) doi:10.2196/38041

KEYWORDS

patient data; comparison; visualization systems; visual analytics; information visualization; cohorts; multiple patients; single patients; time-oriented data

Introduction

Overview

The digitization of health care processes has led to large volumes of digitized patient data, enabling new scenarios for data analytics and visualization. In addition to other forms of data representation, visual representation is becoming increasingly important for describing and analyzing data, as well as for drawing conclusions from data and making decisions based on them. The forms of visual representation are as diverse as the data, stemming from all areas of the health care system, such as the care of patients in various subareas of inpatient medicine, such as internal medicine or surgery, emergency and intensive care, and outpatient medicine. The presentation of individual patient data is as important as the presentation of the aggregated data of groups of individuals with certain characteristics. Thus, it has become increasingly important to present individual patient cases in such a way that they are comparable with each other or with cohorts. This goal becomes even more tangible as visualization systems enable the visual analysis of complex, high-dimensional, and heterogeneous data for different objectives.

Although visualization systems for electronic medical record analysis have been developed for decades, most health care information systems still lack basic information visualization concepts. However, visual analysis and delivery of data in the form of visualizations are of great importance in health care, as such forms of representation can reduce errors and improve care [1] and can also help provide new insights into long-term disease trajectories. Information visualization and visual analytics also address the complexity of long-term, time-oriented patient data by reducing the inherent complexity and facilitating a focus on underlying and hidden patterns [2]. Visualization techniques for temporal data enable clinicians to quickly identify relevant trends in patient health records. Visual comparison techniques help clinicians look for differences between a particular patient's data and his or her group, allowing them to identify, for example, whether treatment needs to be adjusted. In a research context, exploratory data visualization for hypothesis generation is a well-established approach to cohort analysis [3]. By using appropriate interactive visualization techniques, both established and emerging, clinicians and researchers can effectively and efficiently detect patterns, explore relationships, and identify anomalies.

Visualization of time-oriented data is a well-researched area of information visualization across diverse domains, such as finance, the environment, and life sciences. A book by Aigner et al [4] reports on 101 different visualization techniques for time-oriented data and is, to the best of our knowledge, the most detailed review in this area. The research and design of information visualization is a user-centric area and has led to frameworks proposing a classification for a what-why-how differentiation [5]. Munzner [6] suggested a general approach for task-, data-, and user-driven visualization design, which is currently a widespread method in the visualization community and can be applied to time-oriented data [7]. On the basis of the given aspects of data and time (what), as well as user objectives and associated tasks (why), different approaches to visualization techniques (how) have been described [4].

With the increasing availability of clinical patient data for secondary use in clinical research, new opportunities for longitudinal studies and data analyses are emerging. Existing studies have captured time-oriented data visualization in health care [4,8]. However, these reviews do not specifically focus on comparing individuals with other individuals or cohorts. However, the largest and most important task associated with all available data is comparison, for example, with earlier periods of a patient's journey, other similar multiple patients, or a cohort. We anticipate that this task will become increasingly important and diverse in the future.

However, there appears to be a gap in research specifically related to these visual comparison tasks. Rind et al [9] identified this as an open challenge. In their systematic review, they reported on time-oriented data visualization techniques in health care and pointed out the lack of research on the comparison of a single patient to a group of patients with similar histories. For this reason, we specifically address the visualization techniques used to compare single patients with multiple patients or with a cohort and report on the differences and gaps in design for single-patient, multiple-patient, and whole-cohort visualization. Although information visualization and visual analytics are well-established fields and their application in the medical field has been explored for decades, the use of interactive visualizations for the analysis of patients and their cohorts is still a very active area of research. Therefore, we have gathered works from both scientific fields, the medical informatics and visualization community, to provide a comprehensive overview of the state of the art.

This review aims to answer the following research questions (RQs):

- RQ1: Which visualization techniques are used to compare time-oriented patient data with their cohort data?
- RQ2: What visual analysis objectives and tasks are being supported?
- RQ3: What are the characteristics of the visualization systems and applications?

The goal of this study was to provide an overview of visualization techniques for time-oriented data in health care, which support patient comparison. More specifically, we systematically collected literature and report on the interactive visualization techniques that support the comparison of time-oriented data sets of a single patient with those of multiple patients or their cohorts and summarized the use of these techniques. The visualization systems are described according to their medical characteristics, data type categories, and further relevant visualization aspects of such interactions.

Background

Visually Analyzing Data With Information Visualization and Visual Analytics

Historically, the field of visualization research has been divided into 3 subfields: scientific visualization, information visualization, and visual analytics. Although this division is currently sometimes considered too arbitrary and outdated, it helps to structure different techniques and applications. Scientific visualization deals with data that have an inherent spatial reference, such as volume data from medical imaging or atomic coordinates in a molecule. The terms “information visualization” and “visual analytics” are often used synonymously, although they are not synonymous. However, the division is often less clear: *information visualization* represents abstract data in a visual context and expresses patterns or trends that are inherent to the data (using mostly 2D visualization methods, eg, line or bar charts). Information visualizations are often interactive, enabling the manipulation of data or the visualization for in-depth analysis. *Visual analytics* also represents data in an interactive visual context but further supports the discovery and identification process of patterns and trends by combining automated analysis with interactive visualizations; that is, visual analytics systems use information visualization methods to communicate data. Moreover, additional aids for facilitating the understanding of rather complex data are provided and, therefore, are able to support decision-making. The term *visual analytics* was originally coined in 2005 in the context of complex data analysis systems for homeland security [10], where it was already described broadly as “the science of analytical reasoning facilitated by interactive visual interfaces.” Currently, the visual analytics approach is used in many different application areas, ranging from security over software analytics to biology, medicine, and health [11]. Often but not necessarily, it applies machine learning methods to support data analysis.

As it is often unclear whether a system should be classified as interactive information visualization or visual analytics, the reviewed visualization systems covered both areas. Especially in the health sector, time-oriented data visualizations play an important role, both on the level of individual or multiple patients and on the level of entire populations or cohorts; therefore, they are important subjects of research in information visualization and visual analytics. Although existing reviews investigate visual analytic methods and techniques in public health and report on techniques from an epidemiological point of view [8], or focus on visual analytic methods and techniques applied to public health and health services research [12], these review visualizations for populations do not specifically address the visual analysis of single patients.

Investigating Patients: Individual Patients, Multiple Patients, and Cohorts

Historically, the first cornerstones in the field of exploring data visualizations of individual patients were laid in *LifeLines* by Plaisant et al [12]. Numerous visualization systems for electronic patient records or their data analysis have been developed since, such as Knowledge-based Navigation of Abstractions for Visualization and Explanation (KNAVE) [13], KNAVE-II [14], Visualization of Time-Oriented Records [15], LifeLines 2 [16], EventFlow [17], and CareCruiser [18] to name a few. Tools and concepts supporting the visual analysis of patient progression and cohort comparisons are still under active investigation. A recent example is a visual analytics approach that uses dimensionality reduction to summarize and compare individual participants. This method was used to transform intensive care unit data from a controlled animal experiment into 2D curves representing the changing status of participants, with the possibility of characterizing the ensembles of the participants [19]. Another recent study [20] investigated the visual analysis of event sequences in the context of several topics, of which health care constitutes only a minor portion. Research on the applicability of the other approaches to the health care domain is not covered and, thus, constitutes an avenue for future research.

Existing systematic reviews report on the prevalence of electronic health record (EHR) visualization techniques for individual patients and multiple patients [9,21,22]. Most such visualization systems support the task of analyzing either a single patient or multiple patients. Depending on the context and goal of the analysis, multiple patients with increasing numbers can build up to a cohort. Time-oriented patient data comprise event sequences of different data types, which may be categorized (eg, numeric outcomes to categories) or aggregated in time. The same holds for multiple patients. However, time-oriented cohort data (as in epidemiology) differ in that abstract characteristics such as life expectancy or self-reported outcome measures, for example, pain scales, are used for analyses. These data are often reported, for example, as a calculated mean or median across a group of individuals at specified time points.

Comparing Time-Oriented Patient Data

Comparison is a widely supported task in interactive visualization systems [23]. When visually analyzing patient data, the task of comparison is a common part of the process, ranging from comparing information about a single patient to comparing treatment responses at different times and comparing patients in a cohort.

Beyond the context of clinical research and patient care, an increasing number of patients want to manage their own EHRs, analyze their disease progression, and compare it with similar patients, similar to the web-based platform *patientsLikeMe* [24].

However, comparison is not a single clearly defined task but a range of tasks [23]. Brehmer and Munzner [5] specified 3 tasks (or scopes) for the user goal to query a specific target: "identify, compare, and summarize." This is to query within the scope of a single target (identify), multiple targets (compare), or a set of targets (summarize).

The visualization of time-oriented data of a single or multiple patients has been widely explored [9], and most techniques used for visualizing the data of a single patient can be applied to multiple patients up to a certain degree. However, the visualization of cohorts is different in terms of how data are aggregated in time and value. For example, cohort data in clinical trials comprise data that are usually provided at specific time points (number of visits or days aligned for a baseline event) and, in most cases, are represented as statistical values (eg, mean or SDs).

Comparisons within single patients or within cohorts may seem trivial as the same visualization technique is applied for each. However, this may not be the case for comparing a single patient with a cohort, as both may be visualized using a different technique.

Thus, visual comparisons can be supported in various ways. However, the visualization of multiple records can produce visual complexity when visualizing an excessive number of patient records. Similar to Munzner [6], the survey by Gleicher et al [23] emphasizes the exploration of designs for the information visualization of complex data objects, such as graphs, tabular data, and surfaces, and proposes a general taxonomy of visual designs for comparison. Both works differentiate between 3 types of visual comparisons, namely, juxtaposition (or separation), superposition (overlay), and explicit encoding (explicit representation of the relationships), as well as a combination of these. Juxtapositioning means displaying 2 elements that are the subject of the comparison next to each other, whereas superposition means showing them on top of each other in the same view. Gleicher et al [23] found that comparison tasks became more difficult with more complex data objects and when more objects are to be compared, whereas abstracting the data *before* the comparison can simplify the task.

Methods

Protocol and Registration

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) approach. We drafted the protocol for our review by following the checklist in the study by Tricco et al [25]. As this scoping review reports primarily on visualization techniques rather than on the outcomes of medical studies and PROSPERO (International Prospective Register of Systematic Reviews) does not accept scoping reviews, the protocol has not been registered and published.

Eligibility Criteria

On the basis of the presented objective and RQs, we developed criteria for articles to be eligible for review. Articles need to report on a visualization technique, visualization system, or design study supporting the visual analysis of time-oriented patient data to compare a single patient with multiple patients or a patient cohort.

The inclusion criteria were as follows: (1) articles on visualization techniques of time-oriented patient data; (2) articles on systems, applications, or prototypes to support the visual analysis of time-oriented patient data; (3) implementation of tasks to support the visual analysis of time-oriented patient data; and (4) study of a visualization technique for time-oriented data in which physicians or clinical researchers have undergone a test (or questionnaire).

The exclusion criteria were as follows: (1) articles not in English; (2) articles not focused on abstract time-oriented patient data, for example, medical imaging methods (eg, positron emission tomography, magnetic resonance tomography, functional magnetic resonance imaging, and computed tomography); (3) articles on 3D visualizations supporting surgery, operations, or other medical interventions, for example, augmented or virtual reality applications; and (4) articles on deep learning and other machine learning approaches (using patient data), where visualization is solely used to present the implementation.

Articles focusing on medical imaging methods were excluded as they did not fit the information visualization aspect. Although these use imaging methods and are sometimes called (scientific) visualization, they do not visualize abstract time-oriented patient data.

The criteria were revised during the screening process, and we specified the comparison aspect more strictly to exclude articles on visualizations where comparing single patients to a cohort was not supported, either explicitly or implicitly.

Information Sources

To collect potentially relevant articles, we searched the following publication databases: PubMed, IEEE Xplore, ACM Digital Library, and the Web of Science core collection. We identified 4 major areas that reflected the concepts of our RQs: time, visualization, data, and health care. Starting with these, we drafted the main sets of keywords based on terms and synonyms from the literature. Along with an experienced librarian, we further refined our search strategy. The full search queries for the different databases are provided in Multimedia Appendix 1. The search was performed on July 2, 2020. The search results were imported into Citavi reference management software. Duplicates were removed after import.

Search

The initial search strategy was developed by a librarian from the library of the Medical Faculty Mannheim of the University of Heidelberg and was aimed at searching only titles and abstracts. It contained the 4 aspects mentioned earlier: time, visualization, data, and health care. The search strings were reviewed and improved through 3 iterations, and 4 of the

reviewers approved the final search strategy. The final search string for PubMed is shown in Textbox 1.

Details of searches on the other databases can be found in Multimedia Appendix 1.

In addition to the database searches, we identified the following reviews regarding the visualization of time-oriented health care data: West et al [21], Preim and Lawonn [8], and Aigner et al [4].

We examined the reference lists of these reviews and identified 13 articles that we considered a fit for our search but were not included in our search results. To address the search for potential gray literature, we included articles from IEEE VIS annual meetings and workshops on time or sequence visualizations.

Textbox 1. Search aspects and PubMed search string.

Time

- (“temporal data”[tiab] OR “temporal sequence*”[tiab] OR “temporal pattern*”[tiab] OR “temporal abstraction*”[tiab] OR “temporal event*”[tiab] OR “time sequence*”[tiab] OR “time series”[tiab] OR “time period*”[tiab] OR “time frame*”[tiab] OR “timeframe*”[tiab] OR timeline*[tiab] OR time-oriented[tiab] OR (“time”[tiab] AND “events”[tiab])) AND

Visualization

- (visuali*[tiab] OR “visual analy*”[tiab]) AND

Data

- (data[tiab] OR information[tiab]) AND

Health care

- (patient[tiab] OR patients[tiab] OR “health care”[tiab] OR health care[tiab] OR cohort*[tiab] OR “electronic health record*”[tiab])

Selection of Sources of Evidence

In the first screening step, 16 reviewers working in groups of 2 independently screened titles and abstracts for eligibility. In this first screening, we only focused on visualizations of time-oriented patient data and did not include criteria for the comparison of a single patient with a cohort. Disagreements were resolved through discussion and consensus of a third reviewer.

In accordance with our objectives, we discussed the results of the screening and continued to perform a second screening step to apply the criteria of single-to-multiple or cohort comparisons.

The titles, abstracts, and full texts of the remaining articles were skimmed for eligibility for an in-depth full-text analysis. The remaining publications were used for data extraction.

Data-Charting and Extraction Process

For the data extraction, a data-charting form was developed and refined throughout several iterations. The initial form included several categories and abstractions for meta-information, medical context, data, and visualization aspects.

The form was tested by 4 reviewers by applying it to 2 randomly selected articles, of which 1 was assigned to each reviewer. We discussed our findings for corrections and reconciliations throughout the iterative process and released the final version of the form.

Data Items

For each of the included articles, we specifically focused on four major aspects: (1) *meta-information* of the article (authors, year, and digital object identifier), (2) *medical characteristics* (disease, medical context, and medical objective), (3) *data type categories* (type of medical data, data type, temporality, temporal spread, and availability of data set), and (4) *visualization aspects* (visualization technique, tasks, interactions, comparison, and evaluation). We extracted several data items for each aspect.

The meta-information was collected from the respective literature databases. It requires no further categorization but can be used to sort and compare publications, for example, by author or year and venue of publication.

The section on *medical characteristics* comprises Medical Subject Headings (MeSH) terms for the diseases and medical objectives. The medical context was clinical research, clinical care, or both. For the extraction and grouping of medical data types, we used the following categories: encounter (or transfers or movements), diagnosis, procedure, laboratory results, medication, cardiology findings, activity, condition, clinical note, treatment plan, tumor severity, survival, Framingham Risk Score, and patient-reported outcome. The categories were based on hierarchically high-ranked concepts of clinical terminology in the MeSH thesaurus and their frequency in the included studies.

The data *type categories* were further distinguished as qualitative, quantitative, categorical, and free-text. Data temporality was determined for the time primitives (single time points, time intervals, or both) and temporal arrangement as either sequential or cyclic. The temporal spread was extracted as short (from hours up to a few days), long (longer than a few days), and short to long (from hours to several days). Data availability was either described as yes (including restricted availability), no (if not available), or not applicable if no further information was given.

In the context of *visualization*, we applied Visual Vocabulary [26] to the visualization techniques found in the studies. Visual

Vocabulary aims to improve chart literacy for people outside the visualization research community. This visual overview classifies visualization techniques by their main objective and structures them into 9 categories such as part to whole or correlation. The category of *change over time* is particularly relevant to our investigation of temporal patient data and contains techniques such as line charts, calendar heatmaps, or *Priestley timelines*. The latter shows sequential and parallel events on a temporal x-axis, is similar if not synonymous to Gantt and span charts, and is often simply labeled as "event timeline." Although Visual Vocabulary is not a standardized taxonomy, it is used in both academia and practice. Other reviews of visualization techniques [27] used the taxonomy proposed by Borkin et al [28], which is a mix of basic graphs, data, and task-oriented categories but does not include time as a specific category. Wilke [29] discussed temporal data visualizations but did not include them in his Directory of Visualizations.

Visualization systems are intended to support a wide set of tasks ranging from simple ones such as finding the laboratory value of a specific patient at a given date to more complex tasks such as comparing the progression of multiple characteristics of all patients within a cohort. To discuss the similarities and differences of such diverse tasks, different frameworks of abstract task descriptions have been proposed.

The extraction of tasks in our review was based on the widely used taxonomy for task abstractions by Brehmer and Munzner [5], whereas the extraction of actions and targets relied on the taxonomy by Munzner [6].

At the top level of the taxonomy by Munzner [6], visualization systems can be categorized according to the user objectives and associated tasks—why users use visualization techniques in terms of actions and targets. Actions can be classified as *analyze*, *search*, and *query*, and these can be further split into subcategories. Regarding visualization systems of the category of *analyze*, a distinction can be made among, for example, offering data analysis for viewing, understanding information, and creating new information. Consuming information includes the discovery of new insights based on visualized data (*Analyze: Consume: Discover*), as well as using the visualization for presenting insights to others (*Analyze: Consume: Present*) [6].

We opted not to use the health data–specific task taxonomy by Theis et al [30] as it is designed to capture tasks from the perspective of patients. The data-driven taxonomy by Rostamzadeh et al [31] provides a framework for activities and tasks at different levels of granularity (activities, subactivities, tasks, and subtasks) and proposes 3 major categories: interpretation, monitoring, and prediction. However, comparison is not explicitly defined as a task but rather mentioned as an inherent task between interpretation (overview: visually compare) and prediction (recognize: similarity). Consequently, we considered this taxonomy unsuitable for the collection of comparison tasks. Therefore, we applied the taxonomy by Gleicher et al [23] for comparison.

We focused on comparison tasks between different types of relationships: single-to-single patient comparison (1-1), single-to-multiple patients comparison (1-n), single-to-cohort comparison (1-1), and cohort-to-cohort comparison (1-1 or 1-n). "Single to single" comparison means that users can compare individual patient data over time with a nominal or target value or with another single patient. "Multiple patients" stands for ≥2 patients with similar traits or characteristics and, in contrast to cohorts, are an ad hoc group (ie, a dynamically selected subset of patients). This includes comparing data, such as time points and time intervals for procedures, diagnosis, laboratory values, and encounters, across patients. The data are often aggregated, as is the case in flow-based or stage-based approaches (eg, Guo et al [32]). By contrast, cohorts are patient collectives in a clinical or academic research setting; that is, cohorts generally include more patients than "multiple patients." These data tend to be 1D for ≥1 group. This could be averaged over the entire cohort.

More details about the items and their concrete sets of attributes are available in Multimedia Appendix 2 [15,16,18,33-51].

Critical Appraisal of Individual Sources of Evidence

We critically appraised the individual sources specifically for the comparison task. Visual comparisons can be made between ≥2 individual patients, between patients and a cohort, and between cohorts. During the charting process, we systematically collected and collectively discussed visualization techniques in studies in which the applicability to the RQs was in doubt. Included articles that mentioned both visualization and comparison but did not suit our RQs, for instance, because of the comparison of 2 visualization systems, were excluded from the collection. For uncertain cases, in which the explicit comparison of single patients against a cohort was not clearly provided, we critically evaluated whether the visualization technique could implicitly or potentially facilitate that objective.

Synthesis of Results

For the synthesis of results, evidence is presented in the form of charts and tables. For the aforementioned data items, we present and justify our selection of terms, schemas, and taxonomies for different relevant attributes to extract. We combined top-down and bottom-up methods in an iterative approach and adapted and refined the terms where necessary.

We aimed to specifically report on the collected visualization techniques and interactions for the comparison task, as well as summarize the disease, medical objective, and the corresponding medical data types.

For the quantitative analysis, we created charts for articles according to the publication year. We used different tools for the analysis, ranging from a simple dashboard tool for preliminary analysis [52] to Jupyter notebooks, using the data analysis library Pandas and the visualization library Altair for exploratory data analysis. The resulting visualizations were also used to inform the qualitative analysis of the selected characteristics of the articles.

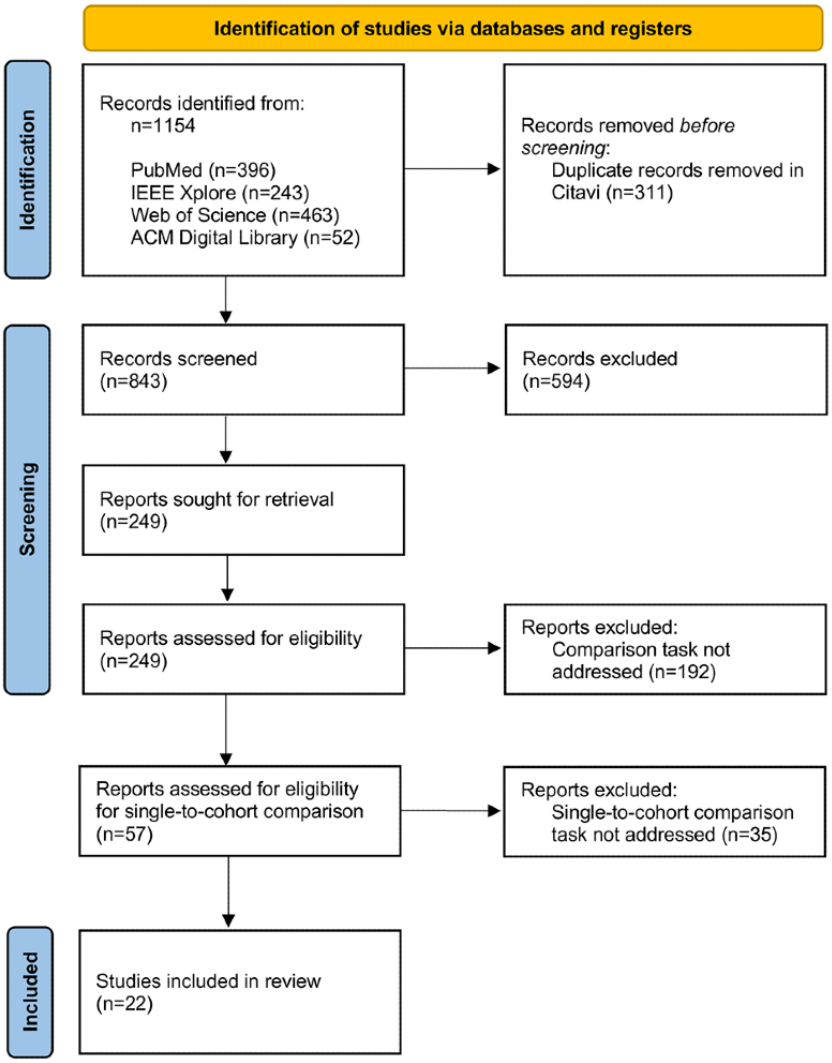
Results

Selection of Sources of Evidence

We identified 1154 articles through individual database searches. Following separate imports into the reference management

software (Citavi) for each database, we removed 26.95% (311/1154) of duplicates electronically. As a first screening step, the titles and abstracts of the 73.05% (843/1154) remaining articles were checked by 16 reviewers, with each article being screened independently by 2 reviewers. Approximately 70.5% (594/843) of papers were excluded based on the study's inclusion and exclusion criteria. In the second screening step, of the 843 articles, we skimmed the full text of 249 (29.5%) articles, of which 192 (22.8%) were excluded as they did not report on the task of comparing patients or cohorts. In the following review step, the full texts of the remaining 57 articles were analyzed in depth for the comparison task, of which 35 (61%) articles were removed as the task of comparing a single patient to multiple patients or a cohort was not provided explicitly, implicitly, or potentially. Of the 843 articles, 22 (2.6%) were included in the synthesis. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram is presented in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the identification, screening, and inclusion of articles.



Characteristics of Sources of Evidence

Overview

We included 22 articles in the scoping review. Most (17/22, 77%) of the included articles were explicitly either single-to-cohort or single-to-multiple comparisons and implemented a comparison visually. The remaining articles (5/22, 23%) were single-to-single (1/5, 20%) or either cohort-to-cohort or multiple-to-multiple (4/5, 80%) comparisons. We included these 5 articles as the presented techniques, in our opinion, could potentially be applied or extended easily to handle a single-to-cohort or single-to-multiple comparison.

All included articles were published between 2003 and 2020. Of the 22 articles, 8 (36%) were published before the review by Rind et al [9] in 2013.

Medical Context

We looked at the medical setting in which the visualization research took place. Medical context in the corpus was mainly clinical research (13/22, 59%), clinical care only (4/22, 18%), and both areas (5/22, 22%; Table 1).

Table 1. Medical context in the selected articles (N=22).

Medical context	Studies	Articles, n (%)
Clinical care	Atherton et al [45], Klimov and Shahar [15], Wang et al [16], and Borhani et al [38]	4 (18)
Clinical research	Gschwandtner et al [18], Gotz and Wongsuphasawat et al [41], Stubbs et al [35], Tao et al [46], Gotz et al [42], Cho et al [47], Browne et al [48], Dabek et al [49], Kamaleswaran et al [40], Gomov et al [39], Wildfire et al [34], Nickerson et al [50], and Polack et al [37]	13 (59)
Clinical research or clinical care	Guo et al [43], Rogers et al [36], van Dortmont et al [33], Magallanes et al [44], and Dahlin et al [51]	5 (22)

Disease

Most of the included articles (9/22, 41%) reported on pathological conditions, signs, and symptoms. The second most frequent diseases mentioned were related to wounds and injuries (2/22, 9%), neoplasms (2/22, 9%), or cardiovascular disease (2/22, 9%).

Medical Objective

Nearly all reviewed articles were found to have "Treatment Outcome" (16/22, 73%) as the primary medical objective. The second most frequent and equally distributed were "Patient Outcome Assessment" (3/22, 14%) and "Disease attributes" (3/22, 14%).

Results of Individual Sources of Evidence

Data

The medical data types identified for visualization and patient comparison contained in the included sources of evidence ranged from the laboratory (13/22, 59%), vital signs (9/22, 41%), and procedures (8/22, 36%) to diagnosis (8/22, 36%). An overview of all extracted data types is provided in Table 2.

The temporal spread of the data was extracted as short (for a couple of hours to less than a few days), long (for more than a few days), and short to long (for data ranging from hours to more than a few days). Most articles reported on either a long (10/22, 45%) or short to long (9/22, 41%) temporal spread, and one of the articles reported on a short (1/22, 4%) spread only. In 4% (1/22) of articles, we could not determine the temporal spread of the data.

Table 2. Medical data types in included articles.

Medical data types	Studies	Articles, n (%)
Laboratory	Atherton et al [45], Klimov and Shahar [15], Wang et al [16], Borhani et al [38], Gschwandtner et al [18], Stubbs et al [35], Gotz and Stavropoulos [42], Browne et al [48], Gomov et al [39] Wildfire et al [34], Guo et al [43], van Dortmont et al [33], and Magallanes et al [44]	13 (59)
Vital signs	Borhani et al [38], Stubbs et al [35], Cho et al [47], Browne et al [48], Gomov et al [39], Wildfire et al [34], Nickerson et al [50], Polack et al [37], and van Dortmont et al [33]	9 (41)
Procedures	Wang et al [16], Gotz and Wongsuphasawat [41], Stubbs et al [35], Tao et al [46], Gomov et al [39], Guo et al [43], Rogers et al [36], van Dortmont et al [33], and Dahlin et al [51]	8 (36)
Diagnosis	Stubbs et al [35], Tao et al [46], Gotz and Stavropoulos [42], Dabek et al [49], Gomov et al, 2017 [39], Guo et al [43], van Dortmont et al [33], and Dahlin et al [51]	8 (36)
Medication	Gotz and Wongsuphasawat [41], Gotz and Stavropoulos [42], Browne et al [48], Gomov et al [39], and Guo et al [43]	5 (23)
Encounters (or transfers or movements)	Wang et al [16], Dabek et al [49], Guo et al [43], van Dortmont et al [33], and Magallanes et al [44]	5 (23)
Patient-reported outcomes (or outcomes)	Atherton et al [45], Gotz and Wongsuphasawat [41], Stubbs et al [35], Nickerson et al [50], and Rogers et al [36]	5 (23)
Cardiology	Stubbs et al [35], Kamaleswaran et al [40], Polack et al [37], van Dortmont et al [33], and Gotz and Wongsuphasawat [41]	5 (23)
Activity	Browne et al [48], Nickerson et al [50], Polack et al [37], and Rogers et al [36]	4 (18)
Conditions	Tao et al [46], Dabek et al [49], and Rogers et al [36]	3 (14)
Clinical notes	Gotz and Wongsuphasawat [41] and van Dortmont et al [33]	2 (9)
Other	Gschwandtner et al [18] and (treatment plans) Dahlin et al [51] (tumor severity and survival)	2 (9)

Visualization Techniques

The visualization system comprised ≥1 visualization technique (mean 2.86, SD 1.36). Most (18/22, 82%) of the articles combined multiple visualization techniques, although some of the used techniques were not explicitly directed at the comparison task and were rather used for auxiliary visualizations or for purposes not related to the comparison. Some articles (4/22, 18%) implemented only 1 visualization technique, whereas some (4/22, 18%) featured more complex visualizations using a combination of up to 5 techniques. Systems provide multiple techniques by showing them either side by side (juxtapositioned, eg, in coordinated multiple views or in a dashboard), overlaid (superpositioned, ie, resulting in combined visualizations), or on different pages within a system (eg, interactively switching between multiple views). We identified all visualization techniques and grouped them according to what they mainly intended to show (see the visualization categories in Figure 2).

In general terms and detached from the restriction of analyzing only the articles with explicit single-to-cohort or multiple comparisons, the major visualization techniques identified in our review are line and *Priestley timeline* charts, histograms, scatterplots, and bar charts.

Out the 22 articles, 17 (77%) included at least one technique to show *change over time*, with 2 (9%) articles [36,37] using 3 techniques from this group and 6 (27%) articles using 2 techniques. Overall, we identified 27 occurrences to visualize the progression of 1 or multiple attributes. The most frequently used were temporal line charts (8/22, 36%) and event timelines (8/22, 36%), followed by columns (4/22, 18%), connected scatterplots (2/22, 9%), and fan charts (2/22, 9%). Techniques used once ranged from calendar heat maps to area charts and candlesticks.

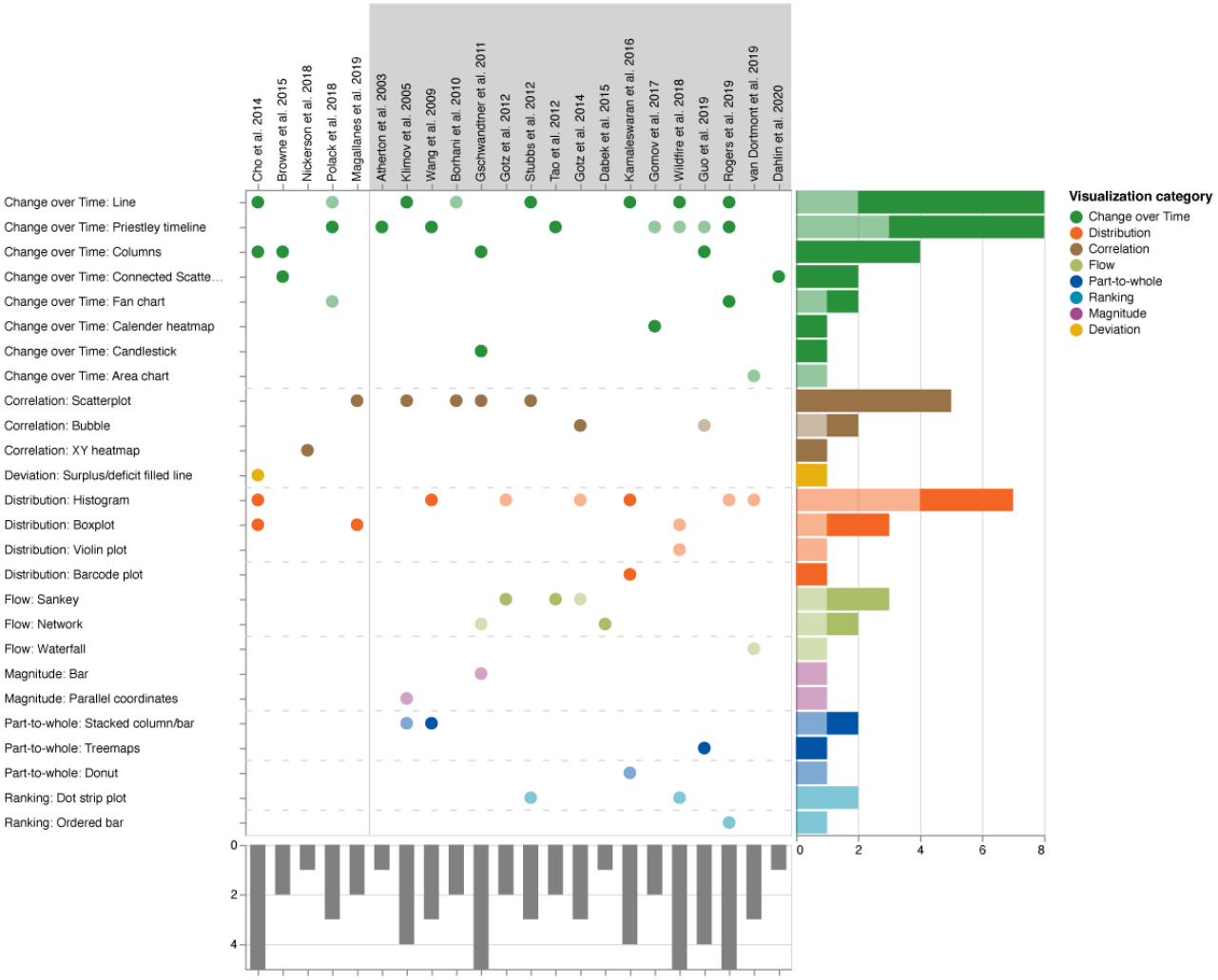
The second largest visualization category was *distribution*, with 41% (9/22) of articles having ≥1 technique from this group. Of the 22 articles, overall, we extracted 12 occurrences of techniques showing the distribution of values: histograms were used in 7 (32%) articles, box plots in 3 (14%), and violin plots and barcode plots in 1 (4%) article each.

The third largest category supports the analysis of *correlation*, with scatterplots (5/22, 23%) and bubble charts (2/22, 9%) being the most applied techniques.

Other techniques used more than once included Sankey charts (3/22, 14%) from the *flow* category, bubble charts, network diagrams, stacked bars, and dot strip plots (2/22 each, 9%) from various other categories.

Using the taxonomy for visual comparison by Gleicher et al [23], we found that most works either applied *juxtaposition* (10/22, 45%), some (4/10, 40%) of which featured an additional *explicit encoding* of the relationship, or *superpositioning* (10/22, 45%), some (3/10, 30%) of which featured *explicit encoding*. Only 9% (2/22) of studies applied both *juxtaposition and superposition*, and only a single study applied an additional *explicit encoding*.

Figure 2. Visualization techniques in the selected articles. Each dot indicates the existence of the technique in a system, with full saturated dots representing the application of the technique for the explicit task of comparison. References within the gray background were identified to support the comparison of multiple patients (single-to-cohort or single-to-multiple). Colors indicate the visualization category, with the bars on the right showing the distribution of the techniques. Bars at the bottom represent the number of techniques identified for each article. Techniques are sorted based on the number of occurrences, and articles are sorted based on the year of publication.



Tasks

User objectives can be characterized by *task* pairs of *actions* and *targets* (compare data items). We have summarized these for the articles in Figure 3.

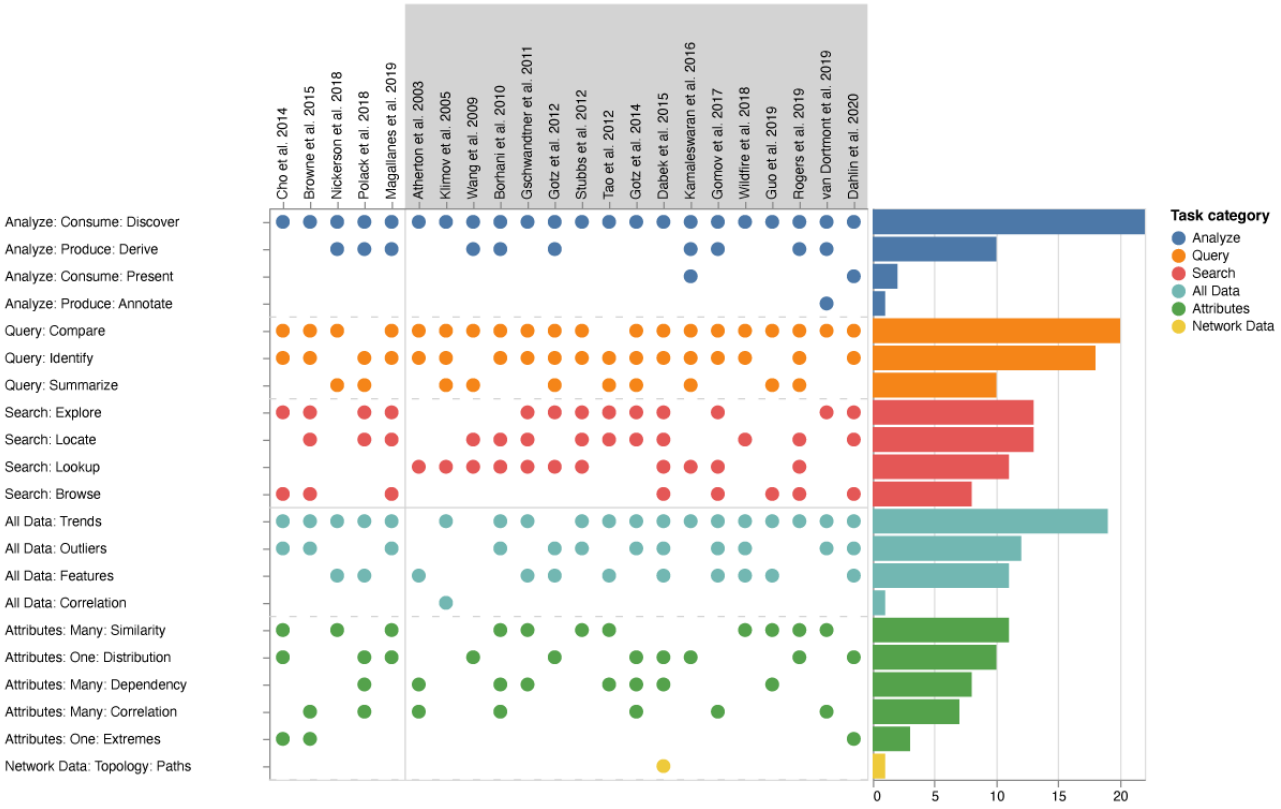
In all the tasks of the reviewed articles, the action of discovering new knowledge in the visualized data is presented (*Analyze: Consume: Discover*). Only 9% (2/22) of articles featured the action of presenting visualized data as the main action (*Analyze: Consume: Present*).

The second most frequent action in this category was *derive*, creating new material from the shown data (*Analyze: Produce: Derive*; 10/22, 45%). Only one of the articles supported annotation (*Analyze: Produce: Annotate*): the ChronoCorrelator supports tagging events with free-form texts that can be used later on, for example, to highlight or filter events for further exploration [33]. In the category of *search*, the actions *locate* (13/22, 59%), finding a known target at an unknown position, and *explore* (13/22, 59%), searching for an unknown target at an unknown position, appeared most frequently. The actions of *lookup* (11/22, 50%), looking for a known element at a known location, and *browse* (8/22, 36%), browsing for ≥1 element without knowing their identity but knowing their characteristics, appeared less frequently. The most frequent action from the category of *query* was *compare* (*Query: Compare*; 20/22, 91%), comparing multiple targets, a result we expected because of our RQ, followed closely by the task of *identify* (*Query: Identify*; 18/22, 82%). The least frequent was the action of *summarize* (*Query: Summarize*; 10/22, 45%).

The targets of the actions were heterogeneous, with a notable exception being *All Data: Trends* (19/22, 86%), which appeared more frequently than others. Targeting outliers (*All Data: Outliers*) and features (*All Data: Features*) was part of 54% (12/22) and 50% (11/22) of articles, respectively. However, correlation as a target (*All Data: Correlation*) was only featured once (1/22, 4%). Actions can target ≥1 attribute of data. The most frequent attribute from the category many was similarity (*Attributes: Many: Similarity*; 11/22, 50%), followed by dependency (*Attributes: Many: Dependency*; 8/22, 36%) and correlation (*Attributes: Many: Correlation*; 7/22, 31%).

The targeting of a single attribute occurred less frequently with distribution (*Attributes: One: Distribution*; 10/22, 45%), appearing far more often than extremes (*Attributes: One: Extremes*; 3/22, 14%).

Figure 3. Identified tasks (actions and targets) in the included articles. The plot shows the tasks as actions (analyze, query, and search) and targets (all data, attributes, and network data) that could be completed by visualization systems presented in the articles in our selection. For the categorization of tasks, we used the taxonomy by Munzner [6]. Gray backgrounds indicate articles where patients (single-to-cohort or single-to-multiple) could be compared. Bars on the right-hand side represent the number of articles that used the displayed task category.



Interaction Techniques

Although interactions offer a way of facilitating a more explorative method of data analysis, more than one-quarter of the articles of interest (6/22, 27%) did not offer any interaction ("No interaction" in the middle of Figure 4) regarding the main task—the comparison of single patients with multiple patients or cohorts.

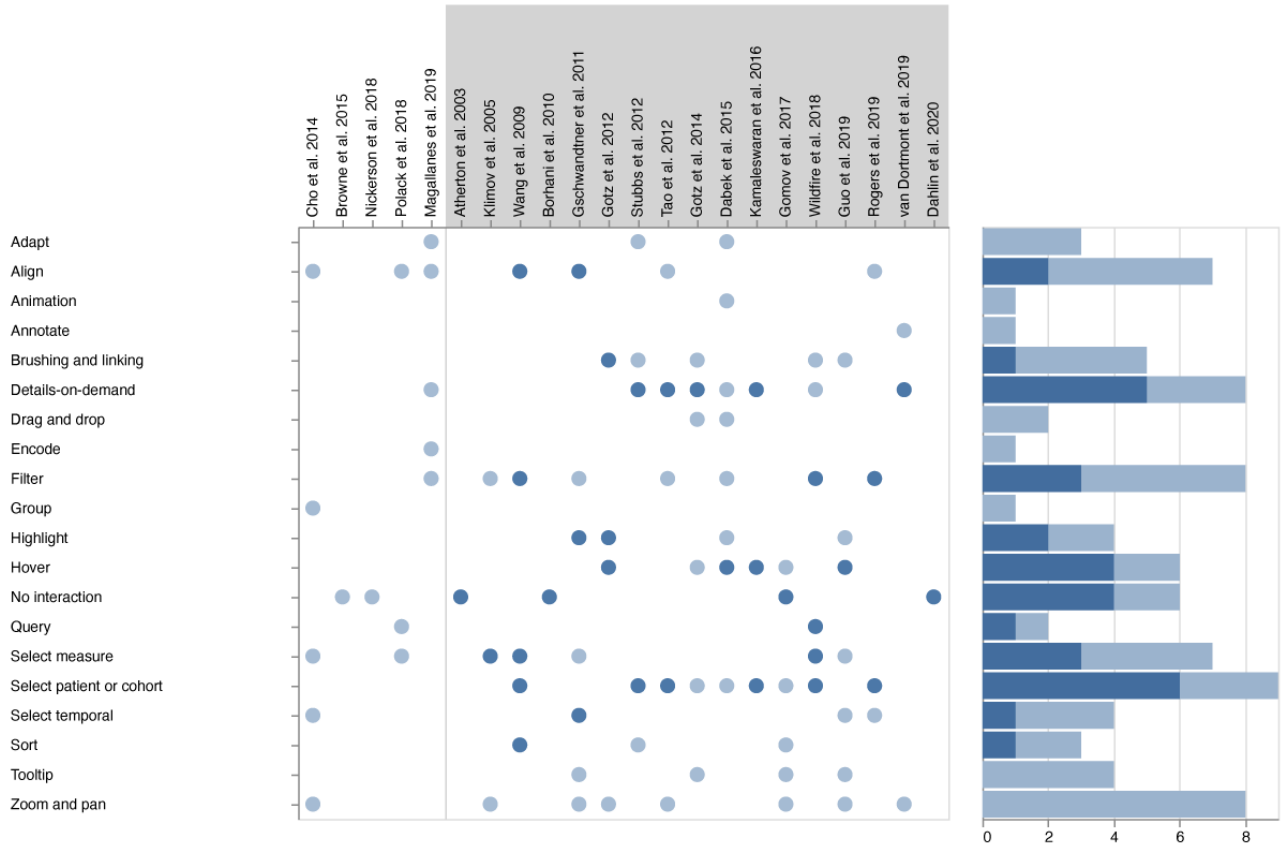
The most frequent interaction possibility found in 41% (9/22) of the articles was the interactive selection of the individual patient or the composition of the cohort. In addition, approximately half of the articles (10/22, 45%) offered an interactive way of showing additional information (details on demand; 8/22, 36%), using hover (6/22, 27%), highlighting (4/22, 18%), or other techniques. The other most commonly used interactions were align (7/22, 32%), filter (8/22, 36%), select measure (7/22, 32%), and zoom and pan (8/22, 36%).

In most studies in scope comparing a single patient with a cohort or multiple patients (17/22, 77%), we identified the selection of a patient and the definition of a cohort as a key interaction technique (*select patient or cohort*; 6/17, 35%).

This was followed by *details on demand* (5/17, 29%) and *hovering* (4/17, 23%). Only 23% (4/17) of these studies did not use any interaction technique for the comparison task.

Depending on the objective of the visualization system to explicitly explore, analyze, and compare patient data, up to 4 interaction techniques are applied directly to support the task for single-to-single or single-to-cohort comparisons [34]. Approximately half of the included articles (9/22, 40%) supported only a single interaction technique for the main task of comparison, whereas some (8/22, 36%) articles combined ≥2 interaction techniques to support the comparison of data elements.

Figure 4. Interaction techniques identified in all included articles. Dark blue dots indicate the explicit application of interaction for the task of comparison. Light blue dots indicate the existence of interactions in the system. References within a gray background were identified to support the comparison of multiple patients (single-to-cohort or single-to-multiple).



Individual Results of Visualization Techniques for Comparisons

Overview

In this section, the individual results of some of the included articles are presented. The most frequent visualization techniques are shown, and a more detailed analysis is provided for line charts, Priestley timelines, scatterplots, and histograms.

For the visual comparison of time-oriented patient data from the category of change over time, line charts and Priestley timelines were used most frequently. The second most commonly used visualization categories were distribution and correlation, with histograms and scatterplots. Depending on the complexity of the presented visualization system, between 1 and 5 visualization techniques were used to visually support a comparison (Figure 2). Table 3 summarizes the visualization techniques used for comparison, the supported combinations of patient entities, and the used visual comparison approach.

Table 3. Overview of all articles containing visualization techniques for comparison.

Author	Visualization technique supporting comparisons	Type of comparison	Visual comparison by Gleicher et al [23]
Kamaleswaran et al [40][a]	• Distribution: barcode plot • Change over time: line • Distribution: histogram	s-s[b], s-m[c], and s-c[d]	• s-s: superposition+juxtaposition • s-m: superposition+juxtaposition • s-c: superposition+juxtaposition
Gomov et al [39][a]	• Change over time: calendar heat map	s-s, s-m, and s-c	• s-s: juxtaposition • s-m: juxtaposition • s-c: juxtaposition+explicit encoding
Atherton et al [45][a]	• Change over time: Priestley timeline	s-s and s-m	• s-s: juxtaposition • s-m: juxtaposition
Gschwandtner et al [18][a]	• Correlation: scatterplot • Change over time: columns • Change over time: candlestick	s-s and s-m	• s-m: juxtaposition+explicit encoding • s-s: juxtaposition+explicit encoding
Tao et al [46][a]	• Change over time: Priestley timeline • Flow: Sankey	s-s and s-m	• s-s: juxtaposition • s-m: juxtaposition
Guo et al [43][a]	• Change over time: columns • Part to whole: tree maps	s-s and s-m	• s-s: juxtaposition • s-m: juxtaposition
Browne et al [48]	• Change over time: columns • Change over time: connected scatterplot	s-s	• s-s: juxtaposition
Wildfire et al [34][a]	• Change over time: line	s-m	• s-m: superposition
Wang et al [16][a]	• Change over time: Priestley timeline • Distribution: histogram • Part to whole: stacked column or bar	s-c, s-m, m-m[e], s-s, and c-c[f]	• s-c: juxtaposition • s-m: juxtaposition+explicit encoding (additive)
Klimov and Shahar [15][a]	• Change over time: line	s-c and s-m	• s-c: superposition • s-m: superposition
Stubbs et al [35][a]	• Change over time: line	s-c and s-m	• s-c: superposition • s-m: superposition
Gotz and Wongsuphasawat [41][a]	• Flow: Sankey	s-c, c-c, c-m[g]	• c-c: superposition
Borhani et al [38][a]	• Correlation: scatterplot	s-c	• s-c: superposition
van Dortmont et al [33][a]	—[h]	s-c	• s-c: superposition+explicit encoding
Gotz and Stavropoulos [42][a]	• Correlation: bubble	c-c, s-m	• c-c: superposition+explicit encoding (animation)
Dabek et al [49][a]	• Flow: network	c-c and s-c	• c-c: juxtaposition • s-c: juxtaposition
Dahlin et al [51][a]	• Change over time: connected scatterplot	c-c and s-c	• c-c: superposition • s-c: superposition
Cho et al [47]	• Deviation: Surplus or deficit filled line • Change over time: columns • Change over time: line	c-c and m-m	• c-c: juxtaposition+explicit encoding (additive)

Author	Visualization technique supporting comparisons	Type of comparison	Visual comparison by Gleicher et al [23]
Rogers et al [36][a]	• Change over time: line • Change over Time: fan chart • Change over time: Priestley timeline	c-c and s-c	• c-c: juxtaposition+explicit encoding, superposition+explicit encoding • s-c: juxtaposition
Nickerson et al [50]	• Correlation: XY heat map	c-c	• c-c: juxtaposition
Polack et al [37]	• Change over time: Priestley timeline	c-c	• c-c: superposition+explicit encoding (additive)
Magallanes et al [44]	• Distribution: box plot • Correlation: scatterplot	c-c	• c-c: superposition

[a]Contain a comparison of single patients to cohorts or to multiple other patients, as visualized in Figure 2.

[b]s-s: single-to-single.

[c]s-m: single-to-multiple.

[d]s-c: single-to-cohort.

[e]m-m: multiple-to-multiple.

[f]c-c: cohort-to-cohort.

[g]c-m: cohort-to-multiple.

[h]Not available.

Visualizing Change Over Time: Line Charts and Timelines

This section provides a qualitative description of some of the key findings regarding the main RQ. Of the 22 articles, 5 (23%) used line charts for comparison, of which 4 (18%) specifically provided a visualization of the task of comparing single patients to multiple or a cohort of other patients. Few (2/22, 9%) of these studies used the *Priestley timelines* in addition to line charts; therefore, they applied 2 change over time techniques.

The line chart used by Klimov and Shahar [15] demonstrated the visualization of a single concept over time; that is, the visualization of a single raw parameter (eg, carbon dioxide) over time for 1 group of patients (Figure 5, top right). The chart displays the top line for the maximal values, the bottom line for the minimal values, and the wide line (thick line) in the middle for the average values in the selected group of patients. The selected patient is displayed as an additional line, which, among other techniques, facilitates the comparison of a single patient with the cohort for this single parameter over time.

In Sim-TwentyFive by Stubbs et al [35], multiple multiline charts (arranged as small multiples) displayed various patient parameters for multiple similar patients (Figure 5, center right). Similar to the study by Klimov and Shahar [15], color coding was used to highlight the queried, selected, or most recently selected patient (green, white, or yellow, respectively), which enabled an easy comparison of the lines of interest, whereas unselected patients remained partially transparent black. In addition, aggregate polygons could be superimposed optionally to visualize the cohort mean and SD of a measure.

Similarly, Wildfire et al [34] used a multiline chart to display the development over time of multiple patients for a single selected patient measure (Figure, 5 top left). The time axis could be switched between days (starting at the baseline event of a study) or visits (the number of visits in a study). At the end of each multiline chart, a box plot representation helped compare the single selected patient line with the overall value across the cohort.

The line chart of Rogers et al [36] allowed a multitude of interactions, ranging from aggregation to normalization, serving primarily to show the development of self-reported patient outcomes over time for different cohorts and individual patients (Figure 5, bottom right). Individual patient scores could be viewed in a multiline chart, which enabled a comparison between patients. Color coding based on the calculated quartiles across the cohort could also enable individual patient comparisons with the cohort.

Composer by Rogers et al [36] and SafetyExplorer by Wildfire et al [34] used *Priestley timelines* in addition to line charts. Composer shows Patient-Reported Outcomes Measurement Information System scores over time as line charts and patient procedure code history in a *Priestley timeline*. The 2 visualizations (Figure 5, right) are time aligned with new time range selections reflected in both views. One or multiple patients can be selected in the nonaggregated line chart whose procedure code histories are shown in the *Priestly timeline*. In contrast, SafetyExplorer provides a line chart and event timeline on separate pages rather than in a coordinated manner. The suite provides all views as components.

Chronodes by Polack et al [37] explicitly showed only the use of Priestly timelines in the context of a cohort-to-cohort comparison. In detail, they used event glyphs called “kebabs” to show the occurrences of specific, differing event sequences preceding or following single or multiple shared sequences of events, called “focal events.” As the relative frequency of the preceding or following sequence was shown, this allowed the user to compare cohorts (ie, groups sharing the same focal event).

The aforementioned articles used these techniques for the task of comparing patients, and several articles also featured the same techniques but for tasks other than comparison.

For example, Borhani et al [38] also used line charts to display individual patient parameters over time but not directly for comparison. Gomov et al [39], among others, also used the Priestly timeline but only to visualize additional data, such as procedures, medication, or infections, on a per-patient basis.

Figure 5. Examples showing line charts as the primary visualization technique for change over time. Use of a line chart to display the cohort (mean, maximum, and minimum boundaries) and a single selected patient (top left) (reproduced from Klimov and Shahar [15], an Open Access article). Use of a line chart to display multiple patients (unselected, selected, and queried individuals) (center left) (reproduced from Stubbs et al [35], an Open Access article). Use of a line chart to display and compare 2 cohorts (mean and quantiles) over time (bottom left). In addition, selected single patients are displayed below as Priestley timelines (reproduced from Rogers et al [36], an Open Access article, which is published under Creative Commons Attribution 4.0 International License [53]). Use of a line chart to display multiple patients (top right). In addition, small multiples (line charts) display more parameters for a selected patient (reproduced from Wildefire et al [34], with permission from Springer, conveyed through Copyright Clearance Center, Inc). Another view in the system in shows Priestley-like timelines (dot-stripe) for individual patients (bottom right) (reproduced from Wildefire et al [34], with permission from Springer, conveyed through Copyright Clearance Center, Inc).

Visualizing Distribution: Histograms

In contrast, 1 common visualization technique was used only sparingly (7/22, 32% articles) for the comparison task: the histogram. Of the 7 uses of histograms, 6 (86%) were featured in studies using single-to-cohort or single-to-multiple comparisons (Figure 2). However, only 33% (2/6) of the studies used histograms for the comparison task directly. These 2 studies, by Wang et al [16] and Kamaleswaran et al [40], used the histogram differently.

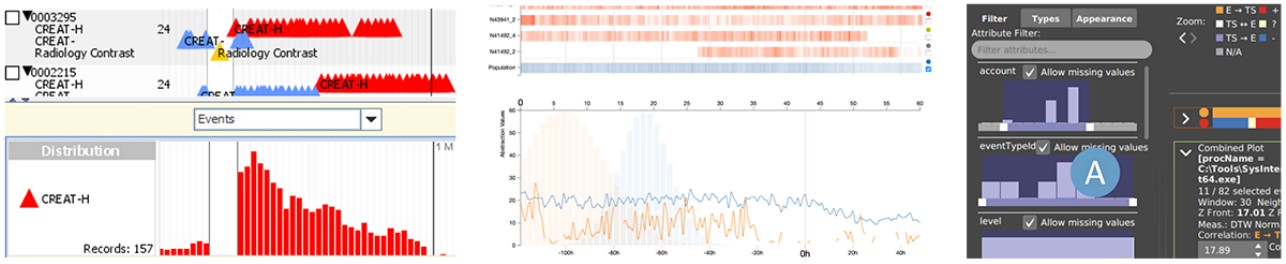
LifeLines2 by Wang et al [16] used an event-aligned timeline in which other types of events could be plotted as a histogram over the complete cohort (Figure 6, left). By using interactions, certain patients with events in specific regions could be selected, and a single patient's event pattern could be directly compared with the general distribution of events, as indicated by the histogram.

Kamaleswaran et al [41] offered a detailed view of their system in which the distribution of a parameter, for example, heart rate variability, over the complete cohort was superimposed with the distribution of measured heart rate variability for a single patient (Figure 6, center). This enabled the direct comparison of a selected patient with the cohort.

The other studies that did not use histograms as a means of direct comparison used them as an auxiliary visualization, displaying additional data. For example, Gotz and Wongsuphasawat [41] used it to show the frequency of the types of interventions or medication in a selected subgroup of patients,

and van Dortmont et al [33] used it as a basis for interactive filtering of the data set (Figure 6, right).

Figure 6. Examples of histograms to display distribution over time. Use of a histogram to display the occurrences of procedures before and after an event (left) (reproduced from Wang et al [16], with permission from IEEE). Use of histograms to display the distribution for a selected measurement of a cohort and an individual (center). Additionally, line charts display the raw data (reproduced from Kamaleswaran et al [40], with permission from the authors). Use of a histogram as an interactive filter (right) (reproduced from van Dortmont et al [33], with permission from the authors).



Visualizing Correlation: Scatterplots

The third largest group contained techniques for exploring and analyzing correlations. The 2 most frequent techniques were scatterplots with 5 occurrences and bubble charts with 2 occurrences, which extended scatterplots by additionally encoding an additional attribute to the size of the marks.

The scatterplots in the reviewed systems enabled comparison through 2 means: showing a connecting line to highlight a single patient (superpositioned) or combining them with an additional technique (juxtapositioned). Approximately 57% (4/7) of the scatterplots showed the derived data by projecting multidimensional data to 2D data [32,38], visualizing correlation as bubble size [42], or calculating a similarity value [35].

Klimov and Shahar [15] visualized the measurement of a parameter over time in a group of patients. Here, the diagram showed the measurements for multiple patients without visual distinction for different patients. When a patient was selected, all measurements were connected by a line. In this way, a single patient could be visually compared with a group of patients.

Borhani et al [38] projected a 4D model onto a 2D plane (Figure 7, top left). The measurements of multiple patients in the "normal" state were shown as a cluster of blue dots. The measurements of the first and last hours of a selected patient were shown within the scatterplot in green and red, respectively. This allowed for quick identification of normal and abnormal measurements of the patient. In addition, the original (ie, nonprojected) measurements of the selected patients were shown in line charts juxtapositioned below.

CareCruiser (Gschwandter et al [18]) showed the parameters over time for multiple patients under investigation (Figure 7, top center). For each patient, a chart visualized the parameter's values over time to view their condition. The time axes were relative to a specified time point; thus, the vertically juxtapositioned charts enabled a direct comparison. Different color-coded bands eased visually identifying relevant events of the patient's development.

The Sim-TwentyFive visualization system [35] enabled querying and comparing episodes and measurements of a selected patient with the 25 most similar other patients (Figure 7, top right). A "cartesian coordinate plot" mapped a calculated score to the x-axis such that the distance to the selected patient indicates their similarity for different measures. Users could switch between different continuous and categorical parameters along the y-axis. The similarity index allowed viewers to compare selected patients with multiple others.

DecisionFlow [42] aggregated event sequences into milestones and intermediate episodes, resulting in visually less complex sequences. DecisionFlow contained a statistical panel with a bubble chart as the main visualization (Figure 7, bottom left). The bubble chart enabled the comparison of events over time and the identification of relevant events for further exploration. Each circle represents an event type, positioned onto 2 axes representing positive or negative support; that is, "the fraction of intermediate episodes in the positive [resp. negative] outcome group containing one or more occurrences of the event type." Its size encoded the correlation, with the additional color showing an odds ratio consistent with all other color codings within the visualization system. The correlation and odds ratios were based on the positive and negative outcome groups. Thus, circles closer to the x-axis represented event types that appeared more often in episodes with a positive outcome and vice versa.

Guo et al [43] presented color-coded circles on a 2D chart to support a visual comparison of event co-occurrence (Figure 7, bottom center). It visualized multiple dimensions on a 2D plane using a dimension reduction technique. This dimension reduction approach resulted in similar events being closer together and dissimilar events being more distant in this chart. Although the used t-distributed stochastic neighbor embedding projection often seemed to show clusters, it is heavily dependent on the chosen parameters of the algorithm. The position of each individual event on the x- and y-axes is semantically ambivalent, and thus, this view is only tangentially related to classic scatterplots.

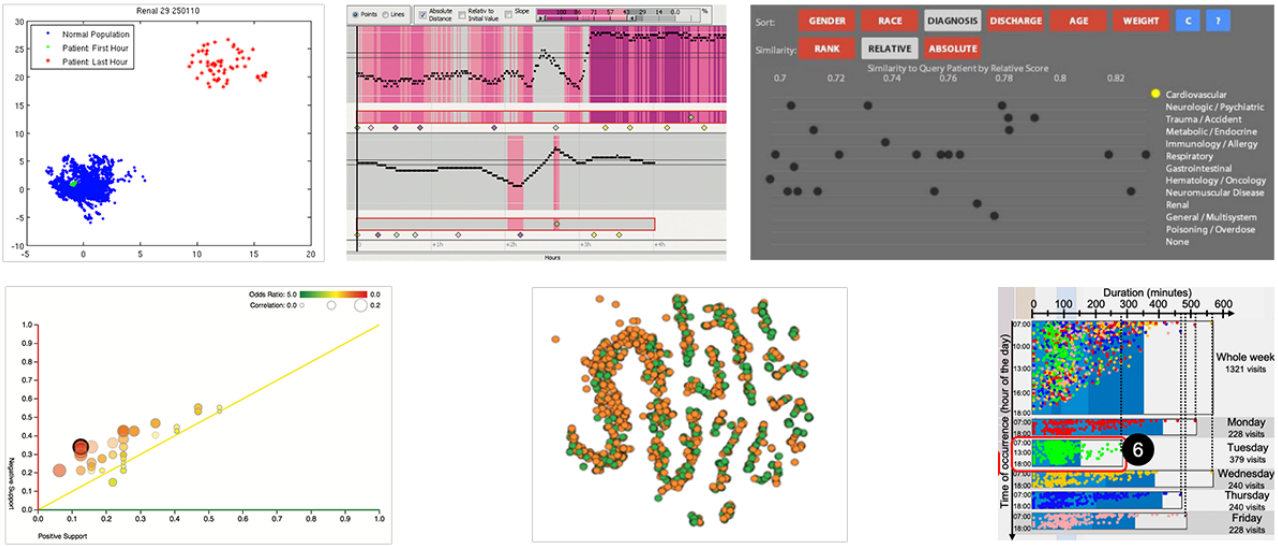
The scatterplot in the study by Magallanes et al [44] enabled the comparison of different weekdays, event sequences, and event occurrences (Figure 7, bottom right). Although it did not facilitate single-to-cohort or single-to-multiple comparisons, it was an unusual approach for visualizing a parameter over time (ie, the occurrence and duration of consultation events). The scatterplot was shown as a superposition of the scatterplots for different patients. This allowed for the quick identification of normal and abnormal measurements. Although the data points can be identified as outliers, the user cannot identify the patient the event occurrence belongs to.

The presented examples of individual results demonstrate common approaches for comparing time-oriented data. Most applied techniques are line charts that show the development of a parameter over time for single or multiple individuals or aggregated for cohorts. Priestley timelines in the presented cases show the periods and mark the start and end of an episode type to be compared but not for directly comparing the quantities.

Bar charts and histograms display the distribution over time and are often used as interactive charts for filtering.

Scatterplots have diverse applications, from simple dots over time to more complex techniques that show correlations between patients and parameters.

Figure 7. Examples of scatterplots to display correlation. Use of a scatterplot to display a 2D projection of values of a cohort and a single patient (top left) (reproduced from Borhani et al [38], with permission from IEEE). Use of a scatterplot to display measurements over time for multiple patients (top center) (reproduced from Gschwandtner et al [18], with permission from IEEE). A scatterplot to show similarity scores across multiple patients (top right) (reproduced from Stubbs et al [35], an Open Access article). Use of a scatterplot (bubble chart) to display the positive and negative outcome contributions of a selected sequence of procedures (bottom left) (reproduced from Gotz et al [42], with permission from the authors). A scatterplot to display a 2D projection of event co-occurrences (bottom center) (reproduced from Guo et al [43], with permission from IEEE). A scatterplot to display occurrences and duration of consultation events (bottom right) (reproduced from Magallanes et al [44], with permission from IEEE).



Discussion

Summary of Evidence

Methodology

The first screening step was conducted by a diverse interdisciplinary team, with contributors having different levels of expertise in visualization research. The last screening and analysis steps were performed by 4 experts from the core team. Although experience and expertise in visualization were advanced, many issues arose during the extraction of data items when applying the different taxonomies. We are aware that in some of our extraction steps, the interpretation of the presented visualizations and application of the corresponding taxonomies may vary. Although we discussed debatable data items, other individuals may obtain different results in some cases.

To provide a systematic overview of the visualization techniques, we investigated different existing taxonomies and classification schemas. We chose Visual Vocabulary as it structures the techniques according to the main objectives. In addition to the task classification by Munzner [6], we collected techniques and their visual analysis objectives and tasks. We found it beneficial to have experts and incorporate publications from both the medical and visualization fields. Through the combination of taxonomies from practice and academia, we were able to collect and review the types of visualizations used for the specific task of comparing temporal patient data. In this way, we could provide an overview of the different visualization techniques and the contexts in which they are used (RQ2).

On a secondary note, we find it worthwhile to highlight how the 2 communities may learn from each other. State-of-the-art reports (STARs) are a major approach to systematically reviewing specific fields in information visualization (McNabb and Laramee [54] and Wang and Laramee [2]). Although they are similarly rigorous in their approach, there is no standardized methodology for collecting and documenting evidence in information visualization reviews. In contrast, STAR articles often use visualizations to summarize their findings. Thus, there might be 2 promising targets for information visualization researchers to build more standardized reviewing and survey procedures and for medical informatics researchers to embrace some of the visual summaries that STAR articles use.

To provide readers with an interactive way of exploring the visualization systems from our scoping review, we created a visual literature browser using the SurVis software [55]. Our tool not only provides a selection of attributes to see the use of specific visualization techniques but also enables cross-filtering to identify systems combining a set of attributes such as medical context, visualization, and patient entities. Our companion tool is available on the web [56].

Medical Characteristics

As synthesized previously, most of the reviewed studies were in the field of clinical research. We assume this to be because

of higher data quality and availability in clinical research, in contrast to data from clinical care, where data are often stored in legacy systems and are not necessarily standardized. A recent survey on EHR visualizations confirmed this assumption; the authors identified 3 challenges impeding the use of EHR data: accessibility, data quality, and interoperability [2].

With respect to the abstracted MeSH terms for coding the diseases, and leaving the generic category of "Pathological Conditions, Signs and Symptoms" aside, we observed a rather wide spread of diseases targeted in the visualization systems. From a medical perspective, this seems to be unexpected, as tumors and cardiovascular diseases are more common. However, from an opportunistic perspective, in selected medical areas, more data are often digitized and easily available, which might result in higher use frequency in medical informatics studies.

This increasing availability of data from primary care facilities enables secondary use in the field of clinical research. In the reviewed studies, we identified the treatment outcome as the major objective of analysis, stemming from both clinical research and clinical care (RQ2). Overall, this emphasizes the need to visually compare changes over time, distributions, and correlations between individuals and their cohorts.

Visualizations

The visual analysis objective most supported in the reviewed systems was to show changes over time. This observation matched our expectations, as the review focused on temporal patient data. Although many medical data have a temporal component, not all visualizations in the medical field focus on time. A scoping review on public health visualizations [12] identified visually analyzing spatial patterns as the most common objective (43.6%), with change over time coming in a distant second, at 14.5%.

To visually investigate correlations, scatterplots and bubble charts were identified as the most common. Here, we noticed that some systems use scatterplots in a nontraditional manner, as they plot a parameter over time on 1 axis [16,19,43]. Although time is a continuous scale and, thus, fits the definition of scatterplots, a more common technique for showing a continuous measure over time is a line chart. When the dots are not sequentially ordered on 1 of the 2 axes but by the values of a nontemporal measure, a connected scatterplot could be used. Both the line chart and connected scatterplot were from the change over time category. We can only assume that the choice of scatterplots (or, perhaps more precisely, scattered dots over time) was because of the design goal of having less cluttered views by omitting the lines. This exemplifies how visualization techniques typically not used for temporal data are used in such ways.

Owing to the nature of single or individual patient data, simple visualization techniques are being used, and the same applies to multiple patients as well, up to a certain degree. In the case of cohorts, which are most often represented as 1D data (aggregated on value and or on time), the same applies and the basic techniques are the most used.

The reported visualization techniques are part of the visualization systems or prototypes of varying maturity levels. Some more advanced and highly interactive systems with a variety of views combined a multitude of techniques, whereas others presented only 1 single and static visualization for 1 objective. We did not evaluate this characteristic and therefore considered the maturity level (complexity of the system, variety of use cases, and tasks) as an interesting parameter for future work. Some articles were simple mock-ups (eg, showing a prototype of a user interface). Other presented articles were edge cases in the sense that the application of the visualization system was primarily developed outside the health care domain, and its application to patient data was shown as a potential use case (eg, ChronoCorrelator showing a use case for analyzing event threads on a server).

Comparison

We identified single patient, multiple patients, and cohort as the entities to visually compare and collected visualization techniques supporting the comparison of any of their combinations. As the reduction of the original search results allowing the comparison of different single patients to the results and the comparison of a single patient to a cohort or multiples was quite noticeable (from 57 to 22), we retained a subset of the studies that would have been dropped at this stage. Thus, these studies (Figure 2) were analyzed in the same manner as the studies explicitly allowing the targeted task. Although these studies might not have been specifically designed to allow the comparison of a single patient with multiple patients or cohorts, the used techniques themselves seemed to be capable of such tasks with little modification. This shows that (1) the visual comparison of a single patient with multiple other patients (single-to-multiple and single-to-cohort) is relatively underdeveloped in comparison with single-to-single or cohort-to-cohort and (2) many existing visualizations purpose-built for the comparison of cohorts among themselves or of a single patient with another individual patient could be adapted to further combinations as well.

By applying our taxonomy for detailed identification of the comparison aspect, we introduced the differentiation between single-to-single, single-to-multiple, single-to-cohort, cohort-to-cohort, multiple-to-multiple, and multiple-to-cohort. Although this differentiation may seem trivial with respect to set theory, it reveals the not directly obvious disruption between showing multiple individuals or a group to be considered the opposing entity of the comparison (Figure 8).

When visualizing patients, we identified the difference between multiple patients and a cohort not in the size of the group, but in the fact that visualizing cohorts requires aggregation of the data beforehand. As shown in our review, this typically goes hand in hand with a different visual representation. For showing a measurement over time, a way of representing the cohort is by visualizing the central tendency (eg, mean) and spread (eg, range) as different lines. An alternative would be to select fitting temporal windows and visualize the spreads of the measurement per time range as box plots.

In Figure 9, we show all possible combinations of visually comparing measurements over time between different patient entities. Figure 9 exemplifies this for line charts, whereas the conceptual space of the different comparison combinations is

agnostic to the used visualization technique. In addition, we only show variations for juxtapositioned versus superpositioned layouts, whereas a wide range of alternate options such as interactions exist. Choosing the appropriate visualization, layout position, and interaction is a major challenge in designing visual analysis systems and requires human-centric development approaches to match the visualizations with the tasks and requirements of users. Overall, although this visual example using line charts provides some initial hints into what might work better than other combinations (eg, superpositioned multiple-to-multiple comparisons seem to be visually overly complex), it is an early exploration of a design space.

Figure 8. Schematic overview of the possible comparisons between different patient entities ranging from single patient to multiple patients and cohorts. Multiple-to-cohort emphasizes the distinction between the visual representation of multiple patients versus an aggregated view of the cohort.

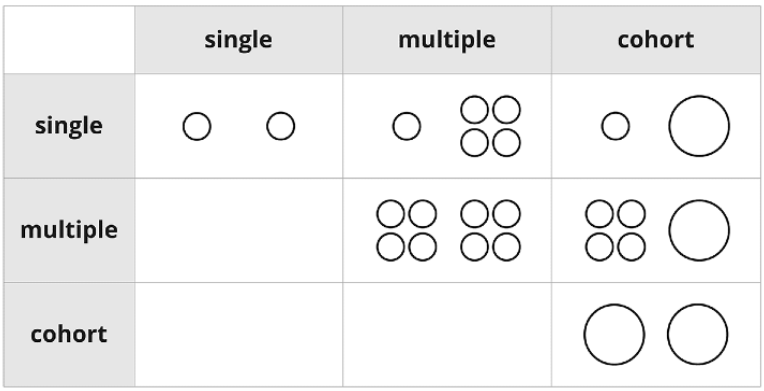
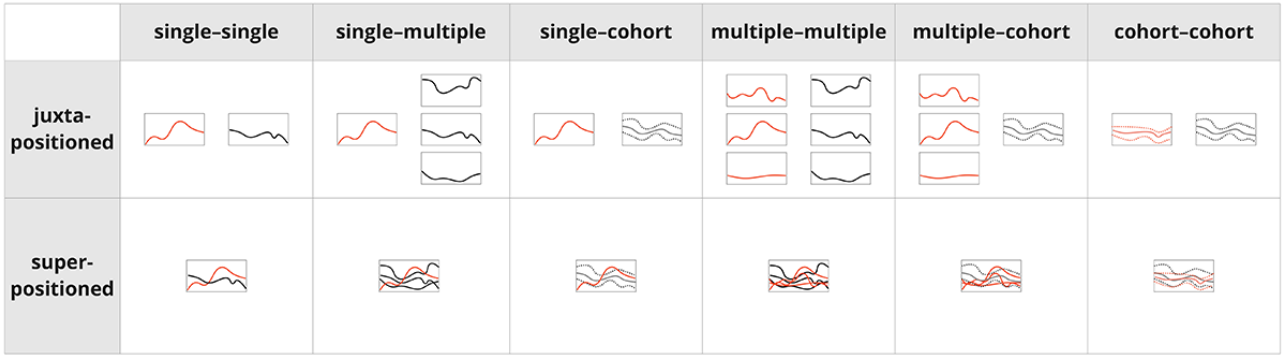


Figure 9. All possible combinations for comparing 1 measurement over time between different patient entities (single patient, multiple patients, and cohorts) in the case of line charts. For single and multiple patients, each line represents 1 measurement, whereas, for cohorts, the chart represents the mean and range. All combinations are shown in juxtapositioned and superpositioned layouts, with the colors supporting legibility in the latter.



Limitations

Although we did not restrict our search to journals and included conference proceedings as these are one of the primary types of publications in the computer science field, we found only a small number of articles that matched all criteria. As the importance of visual analytics in general and visual analytic systems in particular continues to grow, we expected to include more articles from recent years; however, only a few were identified to match our criteria. As described earlier, we iteratively refined our search terms in collaboration with an experienced librarian, and therefore, we assume this to be because of the particular combination of patient-to-cohort comparisons and visualizations focusing on time-oriented data. However, we are aware that we may have missed relevant works; for example, systems that could be primarily cohort visualization tools might also support some detailed highlighting of individual patients without mentioning this explicitly or discussing it in their written report.

We extracted and synthesized a wide set of relevant attributes to summarize the major characteristics of the reviewed studies. However, there is a range of further investigations that we found to be outside the scope of this review. Although we took various specifics of the data into account, we did not evaluate data preparation or data transformation steps alone if they were not an essential aspect of the used visualization technique (such as showing high-dimensional data in a 2D display).

Studies on visual analysis systems usually gather feedback through usability evaluations or demonstrate its applicability through case studies. Although we did not synthesize such attributes in our scoping review, we acknowledge the importance of understanding user feedback to properly assess the usefulness of visualization systems and emphasize the need for further research in this regard.

In addition to the authors and publication year, we restricted the metadata extraction to information about the publication outlet. Analyzing this would allow us to explore the correlations between the extracted attributes and the research area. For instance, one could investigate whether visualization researchers use more complex visualization techniques than researchers in the medical field. This could not be covered in this review, and we did not incorporate it into the analysis.

Furthermore, we did not fully synthesize combinations of the extracted attributes. For instance, it might be insightful to further examine the kinds of interactions that are offered more frequently for specific visualization techniques. The investigation of this explicit combination could lead to a better understanding of which selection techniques for 1 or multiple patients with specific characteristics are appropriate for different aggregated cohort visualizations. Although the articles were analyzed to include visualization of task-specific actions and targets, in this review, it could not be evaluated in further detail whether specific action-target pairs appeared more or less frequently. However, the analysis of these pairs could lead to interesting RQs in the field of visualization research. Our web-based companion tool at [55] provides the first basic possibility of exploring combinations of extracted attributes such as medical diseases and visualization techniques.

Conclusions

Visual analytic systems mitigate the complexity of time-oriented patient data through data analysis and interactive visualizations by facilitating attention to underlying and hidden patterns. In this scoping review, we examined the available literature and identified and clustered visualization techniques that specifically supported the task of comparing time-oriented patient data (RQ1). We collected and reported on the visual analysis objectives and tasks with a specific focus on the range of options to compare individual patients with multiple patients or with a cohort (RQ2). Finally, we surveyed and presented the medical characteristics, data type categories, and interaction techniques of the reviewed visualization systems (RQ3).

As this work is a scoping review, we consider the identified articles and the performed extraction steps as the first step for conducting further research in the form of a more advanced extraction. We found that a small set of publications specifically contained single-to-multiple or single-to-cohort comparison and provided visualizations to support this task. In most cases, we also found that basic visualization techniques such as line charts, event timelines, histograms, or scatterplots were used efficiently. Time-oriented comparisons between a single patient and multiple patients or a cohort are mostly used for laboratory and vital sign parameters, followed by analysis and comparison of procedures and diagnoses. We identified many potentially interesting approaches and deemed many of these techniques to be applicable for a comparison of single patients with multiple patients and cohorts through small adaptations.

We anticipate that we have convincingly argued for the usefulness of visually comparing individual patients with cohorts and encourage researchers to further investigate visualization and interaction techniques for such comparisons. Finally, our review showed the need to systematically review further systems and techniques to propose a proper design space for comparing the temporal data of single, multiple, and cohort patients.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research, Medical Informatics for Research and Care in University Medicine Consortium (01ZZ1801E, 01ZZ1801I, and 01ZZ1801B), and ZIV Consortium (Center for Innovative Care, digital@bw, 42-04HV.MED(18)/25/1). NB, MK, and TN were partially funded by Carl-Zeiss-Stiftung. The authors would like to thank Silke Kühlwein, Kerstin Gierend, Preetha Moorthy, Julia Müller, Alex Liebler, Melanie Börries, Tobias Gradinger, Daniela Zöller, Nico Pfeifer, Daniela Kassahn, and Sylvia Herter for their time and contributions during the early screening steps of this review.

Authors' Contributions

All authors have contributed significantly to this scoping review. JS, TG, MB, and TN conceptualized this review. JS, NB, AV, and TN reviewed the sources of evidence and extracted and validated the data. LW and NS contributed to data acquisition. TG and MB contributed to the extraction and validation of medical aspects. NB and MK contributed to the extraction and validation of visualization aspects. JS and TN performed data analysis and visualization of the extraction results. LW and NS created the companion website. JS, NB, AV, and TN mainly wrote the manuscript, with MK, TG, and MB contributing to the writing and reviewing of the manuscript. All authors revised the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries on literature databases and the number of results.
[DOCX File , 15 KB - jmir_v24i10e38041_app1.docx]

Multimedia Appendix 2
Extraction table with data items for the sources of evidence.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v24i10e38041_app2.xlsx]

References

1. Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. Computer 2013 May;46(5):58-66. [doi: 10.1109/MC.2013.38]
2. Wang Q, Laramee RS. EHR STAR: the state‐of‐the‐art in interactive EHR visualization. Comput Graphics Forum 2022 Feb;41(1):69-105. [doi: 10.1111/cgf.14424]
3. Zhang Z, Gotz D, Perer A. Iterative cohort analysis and exploration. Inf Vis 2014 Mar 19;14(4):289-307. [doi: 10.1177/1473871614526077]
4. Aigner W, Miksch S, Schumann H, Tominski C. Visualization of Time-Oriented Data. London, UK: Springer; 2011.
5. Brehmer M, Munzner T. A multi-level typology of abstract visualization tasks. IEEE Trans Visual Comput Graphics 2013 Dec;19(12):2376-2385. [doi: 10.1109/tvcg.2013.124]
6. Munzner T. Visualization Analysis and Design. Boca Raton, FL, USA: CRC Press; 2014.
7. Miksch S, Aigner W. A matter of time: applying a data–users–tasks design triangle to visual analytics of time-oriented data. Comput Graph 2014 Feb;38:286-290. [doi: 10.1016/j.cag.2013.11.002]
8. Preim B, Lawonn K. A survey of visual analytics for public health. Comput Graphics Forum 2020 Feb;39(1):543-580. [doi: 10.1111/cgf.13891]
9. Rind A, Wang TD, Aigner W, Miksch S, Wongsuphasawat K, Plaisant C, et al. Interactive Information Visualization to Explore and Query Electronic Health Records. Delft, The Netherlands: Now Foundations and Trends; 2011.
10. Thomas JJ, Cook KA. Illuminating the Path: The Research and Development Agenda for Visual Analytics. Richland, WA, USA: National Visualization and Analytics Center; 2005.
11. Keim DA, Mansmann F, Schneidewind J, Thomas J, Ziegler H. Visual analytics: scope and challenges. In: Simoff SJ, Böhlen MH, Mazeika A, editors. Visual Data Mining. Berlin, Germany: Springer; 2008:76-90.
12. Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B. LifeLines: visualizing personal histories. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1996 Presented at: CHI '96; April 13-18, 1996; Vancouver, Canada p. 221-227. [doi: 10.1145/238386.238493]
13. Shahar Y, Cheng C. Knowledge-based visualization and navigation of time-oriented clinical data and their abstractions. In: Proceedings of the 13th European Conference on Artificial Intelligence - Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) Workshop. 1998 Presented at: ECAI '98; August 23-28, 1998; Brighton, UK p. 86-95 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.18.6873&rep=rep1&type=pdf
14. Shahar Y, Goren-Bar D, Boaz D, Tahan G. Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. Artif Intell Med 2006 Oct;38(2):115-135. [doi: 10.1016/j.artmed.2005.03.001] [Medline: 16343873]
15. Klimov D, Shahar Y. A framework for intelligent visualization of multiple time-oriented medical records. AMIA Annu Symp Proc 2005;2005:405-409 [FREE Full text] [Medline: 16779071]
16. Wang TD, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, et al. Temporal summaries: supporting temporal categorical searching, aggregation and comparison. IEEE Trans Vis Comput Graph 2009;15(6):1049-1056. [doi: 10.1109/TVCG.2009.187] [Medline: 19834171]
17. Monroe M, Lan R, Lee H, Plaisant C, Shneiderman B. Temporal event sequence simplification. IEEE Trans Vis Comput Graph 2013 Dec;19(12):2227-2236. [doi: 10.1109/TVCG.2013.200] [Medline: 24051789]
18. Gschwandtner T, Aigner W, Kaiser K, Miksch S, Seyfang A. CareCruiser: exploring and visualizing plans, events, and effects interactively. In: 2011 IEEE Pacific Visualization Symposium. 2011 Presented at: PacificVis '11; March 1-4, 2011; Hong Kong, China p. 43-50. [doi: 10.1109/pacificvis.2011.5742371]
19. Birch N, Schulz C, Peter J, Klingert W, Schenk M, Weiskopf D, et al. Visual analysis of multivariate intensive care surveillance data. In: Eurographics Workshop on Visual Computing for Biology and Medicine. 2020 Presented at: VCBM '20; September 28-October 1, 2020; Tübingen, Germany. [doi: 10.2312/vcbm.20201174]
20. Guo Y, Guo S, Jin Z, Kaul S, Gotz D, Cao N. A survey on visual analysis of event sequence data. IEEE Trans Vis Comput Graph (forthcoming) 2021 Jul 27:1. [doi: 10.1109/TVCG.2021.3100413] [Medline: 34314358]
21. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. J Am Med Inform Assoc 2015 Mar;22(2):330-339 [FREE Full text] [doi: 10.1136/amiajnl-2014-002955] [Medline: 25336597]
22. Rostamzadeh N, Abdullah SS, Sedig K. Visual analytics for electronic health records: a review. Informatics 2021 Feb 23;8(1):12. [doi: 10.3390/informatics8010012]
23. Gleicher M, Albers D, Walker R, Jusufi I, Hansen CD, Roberts JC. Visual comparison for information visualization. Inf Vis 2011 Sep 07;10(4):289-309. [doi: 10.1177/1473871611416549]
24. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. chart-doctor/visual-vocabulary at main. GitHub. 2021 Feb 18. URL: https://github.com/Financial-Times/chart-doctor [accessed 2022-05-23]

27. Tennekes M, Chen M. Design space of origin - destination data visualization. Comput Graphics Forum 2021 Jun 29;40(3):323-334. [doi: 10.1111/cgf.14310]
28. Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, et al. What makes a visualization memorable? IEEE Trans Vis Comput Graph 2013 Dec;19(12):2306-2315. [doi: 10.1109/TVCG.2013.234] [Medline: 24051797]
29. Wilke CO. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Sebastopol, CA, USA: O'Reilly Media; 2019.
30. Theis S, Rasche PW, Bröhl C, Wille M, Mertens A. Task-data taxonomy for health data visualizations: web-based survey with experts and older adults. JMIR Med Inform 2018 Jul 09;6(3):e39 [FREE Full text] [doi: 10.2196/medinform.9394] [Medline: 29986844]
31. Rostamzadeh N, Abdullah SS, Sedig K. Data-driven activities involving electronic health records: an activity and task analysis framework for interactive visualization tools. Multimodal Technol Interact 2020 Mar 03;4(1):7. [doi: 10.3390/mti4010007]
32. Guo S, Xu K, Zhao R, Gotz D, Zha H, Cao N. EventThread: visual summarization and stage analysis of event sequence data. IEEE Trans Vis Comput Graph 2018 Jan;24(1):56-65. [doi: 10.1109/TVCG.2017.2745320] [Medline: 28866586]
33. van Dortmont MA, van den Elzen S, van Wijk JJ. ChronoCorrelator: enriching events with time series. Comput Graphics Forum 2019 Jul 10;38(3):387-399. [doi: 10.1111/cgf.13697]
34. Wildfire J, Bailey R, Krouse RZ, Childress S, Sikora B, Bryant N, et al. The safety explorer suite: interactive safety monitoring for clinical trials. Ther Innov Regul Sci 2018 Nov;52(6):696-700 [FREE Full text] [doi: 10.1177/2168479018754846] [Medline: 29714563]
35. Stubbs B, Kale DC, Das A. Sim•TwentyFive: an interactive visualization system for data-driven decision support. AMIA Annu Symp Proc 2012;2012:891-900 [FREE Full text] [Medline: 23304364]
36. Rogers J, Spina N, Neese A, Hess R, Brodke D, Lex A. Composer-visual cohort analysis of patient outcomes. Appl Clin Inform 2019 Mar;10(2):278-285 [FREE Full text] [doi: 10.1055/s-0039-1687862] [Medline: 31018234]
37. Polack Jr PJ, Chen ST, Kahng M, DE Barbaro K, Basole R, Sharmin M, et al. Chronodes: interactive multifocus exploration of event sequences. ACM Trans Interact Intell Syst 2018 Feb;8(1):2 [FREE Full text] [doi: 10.1145/3152888] [Medline: 29515937]
38. Borhani Y, Fleming S, Clifton DA, Sutherland S, Hills L, Meredith D, et al. Towards a data fusion model for predicting deterioration in dialysis patients. In: Proceedings of the 2010 Computing in Cardiology. 2010 Presented at: CinC '10; September 26-29, 2010; Belfast, UK p. 967-970.
39. Gomov M, Chou JK, Li JK, Sen S, Cho K, Tran N, et al. Aiding infection analysis and diagnosis through temporally-contextualized matrix representations. In: Proceedings of the 2017 IEEE Workshop on Visual Analytics in Healthcare. 2017 Presented at: VAHC '17; October 1, 2017; Phoenix, AZ, USA p. 31-38. [doi: 10.1109/vahc.2017.8387498]
40. Kamaleswaran R, James A, Collins C, McGregor C. CoRAD: visual analytics for cohort analysis. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics. 2016 Presented at: ICHI '16; October 4-7, 2016; Chicago, IL, USA p. 517-526. [doi: 10.1109/ichi.2016.93]
41. Gotz D, Wongsuphasawat K. Interactive intervention analysis. AMIA Annu Symp Proc 2012;2012:274-280 [FREE Full text] [Medline: 23304297]
42. Gotz D, Stavropoulos H. DecisionFlow: visual analytics for high-dimensional temporal event sequence data. IEEE Trans Vis Comput Graph 2014 Dec;20(12):1783-1792. [doi: 10.1109/TVCG.2014.2346682] [Medline: 26356892]
43. Guo S, Jin Z, Gotz D, Du F, Zha H, Cao N. Visual progression analysis of event sequence data. IEEE Trans Vis Comput Graph 2019 Jan;25(1):417-426. [doi: 10.1109/TVCG.2018.2864885] [Medline: 30136953]
44. Magallanes J, van Gemeren L, Wood S, Villa-Uriol MC. Analyzing time attributes in temporal event sequences. In: Proceedings of the 2019 IEEE Visualization Conference. 2019 Presented at: VIS '19; October 20-25, 2019; Vancouver, Canada p. 1-5. [doi: 10.1109/visual.2019.8933770]
45. Atherton PJ, Jasperson B, Nibbe A, Clement-Brown KA, Allmer C, Novotny P, et al. What happened to all the patients? Event charts for summarizing individual patient data and displaying clinically significant changes in quality of life data. Drug Information J 2003 Dec 30;37(1):11-21. [doi: 10.1177/009286150303700104]
46. Tao C, Wongsuphasawat K, Clark K, Plaisant C, Shneiderman B, Chute CG. Towards event sequence representation, reasoning and visualization for EHR data. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. 2012 Presented at: IHI '12; January 28-30, 2012; Miami, FL, USA p. 801-806. [doi: 10.1145/2110363.2110461]
47. Cho M, Kim B, Bae HJ, Seo J. Stroscope: multi-scale visualization of irregularly measured time-series data. IEEE Trans Vis Comput Graph 2014 May;20(5):808-821. [doi: 10.1109/TVCG.2013.2297933] [Medline: 26357300]
48. Browne SH, Behzadi Y, Littlewort G. Let visuals tell the story: medication adherence in patients with type II diabetes captured by a novel ingestion sensor platform. JMIR Mhealth Uhealth 2015 Dec 31;3(4):e108 [FREE Full text] [doi: 10.2196/mhealth.4292] [Medline: 26721413]
49. Dabek F, Chen J, Garbarino A, Caban JJ. Visualization of longitudinal clinical trajectories using a graph-based approach. In: Proceedings of the 2015 Workshop on Visual Analytics in Healthcare. 2015 Oct Presented at: VAHC '15; October 25, 2015; Chicago, IL, USA p. 1-7. [doi: 10.1145/2836034.2836039]

50. Nickerson PV, Baharloo R, Wanigatunga AA, Manini TM, Tighe PJ, Rashidi P. Transition icons for time-series visualization and exploratory analysis. IEEE J Biomed Health Inform 2018 Mar;22(2):623-630 [FREE Full text] [doi: 10.1109/JBHI.2017.2704608] [Medline: 28534797]
51. Dahlin S. Exploring the usefulness of Lexis diagrams for quality improvement. BMC Med Inform Decis Mak 2020 Jan 08;20(1):7 [FREE Full text] [doi: 10.1186/s12911-019-1017-3] [Medline: 31915004]
52. Home of exploratory data analysis. Keshif. URL: https://keshif.me/ [accessed 2022-05-23]
53. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2022-10-17]
54. McNabb L, Laramee RS. Survey of Surveys (SoS) - mapping the landscape of survey papers in information visualization. Computer Graphics Forum 2017 Jul 04;36(3):589-617. [doi: 10.1111/cgf.13212]
55. Beck F, Koch S, Weiskopf D. Visual analysis and dissemination of scientific literature collections with SurVis. IEEE Trans Vis Comput Graph 2016 Jan;22(1):180-189. [doi: 10.1109/TVCG.2015.2467757] [Medline: 26529698]
56. Visualization techniques of time-oriented data for the comparison of single patients to multiple patients or cohorts. GitHub. URL: https://imi-hd.github.io/ComPatiVis/ [accessed 2022-09-12]

Abbreviations

EHR: electronic health record
KNAVE: Knowledge-based Navigation of Abstractions for Visualization and Explanation
MeSH: Medical Subject Headings
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
PROSPERO: International Prospective Register of Systematic Reviews
RQ: research question
STAR: state-of-the-art report

Edited by R Kukafka; submitted 16.03.22; peer-reviewed by H He, T Vaughan; comments to author 14.05.22; revised version received 28.05.22; accepted 28.07.22; published 24.10.22.

Please cite as:
Scheer J, Volkert A, Brich N, Weinert L, Santhanam N, Krone M, Ganslandt T, Boeker M, Nagel T
Visualization Techniques of Time-Oriented Data for the Comparison of Single Patients With Multiple Patients or Cohorts: Scoping Review
J Med Internet Res 2022;24(10):e38041
URL: https://www.jmir.org/2022/10/e38041
doi:10.2196/38041
PMID:36279164

Review

Use of Telemedicine in Pediatric Services for 4 Representative Clinical Conditions: Scoping Review

Genevieve Southgate[1*], BN (Hons), BMedSci, BMBS; Arrash A Yassaee[2,3*], BA (Hons), MBA, MBBS; Matthew J Harmer[1], BSc (Hons), BM, PhD; Helen Livesey[4], BSc (Hons), MBChB; Kate Pryde[1], BMedSci, BMBS; Damian Roland[5,6], BMedSci, BMBS, PhD

[1]Southampton Children's Hospital, Southampton, United Kingdom

[2]Centre for Paediatrics and Child Health, Faculty of Medicine, Imperial College, London, United Kingdom

[3]Huma Therapeutics, London, United Kingdom

[4]Leeds Teaching Hospitals National Health Service Trust, Leeds, United Kingdom

[5]Social Science Applied Healthcare and Improvement Research (SAPPHIRE) Group, Health Sciences, University of Leicester National Health Service Trust, Leicester, United Kingdom

[6]Paediatric Emergency Medicine Leicester Academic (PEMLA) Group, University Hospitals of Leicester National Health Service Trust, Leicester, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Arrash A Yassaee, BA (Hons), MBA, MBBS
Centre for Paediatrics and Child Health
Faculty of Medicine
Imperial College
South Kensington Campus
London, SW7 2AZ
United Kingdom
Phone: 44 (0)20 7589 5111
Email: arrash.yassaee@nhs.net

Abstract

Background: Telemedicine is becoming routine in health care. Postpandemic, a universal return to face-to-face consultations may risk a loss of some of the advantages of telemedicine. However, rapid implementation and adoption without robust evaluation of usability, efficacy, and effectiveness could potentially lead to suboptimal health outcomes and downstream challenges to providers.

Objective: This review assesses telemedicine interventions against international guidance and sufficiency of evidence to support postpandemic utilization in pediatric settings.

Methods: This scoping review was performed following searches on PubMed, Embase, and CINAHL databases on April 15, 2021, and May 31, 2022, and examined studies focused on telemedicine, remote consultation, video call, or remote patient monitoring in children (0-18 years) receiving outpatient care for diabetes, asthma, epilepsy, or renal disease. Exclusion criteria included studies published before 2011 as the technologies used have likely been improved or replaced, studies in adult populations or where it was not possible to disaggregate data for participants younger than 18 years as the focus of the review was on pediatric care, and studies not published in English. Data were extracted by 4 authors, and the data were corroborated by a second reviewer. Studies were examined for feasibility and usability, clinical and process outcomes, and cost-effectiveness.

Results: Of the 3158 studies identified, 56 were suitable for final inclusion and analysis. Data on feasibility or usability of interventions (48 studies) were overwhelmingly positive in support of telemedicine interventions, with common themes including convenience, perceived cost savings, and ease of use. However, use in preference to usual care was rarely explored. Clinical and process outcome data (31 studies) were mostly positive. Across all studies, there was limited measurement of standardized clinical outcomes, although these were more commonly reported in asthma (peak flow) and diabetes (glycated hemoglobin [HbA1c]). Implementation science data generally supported cost-effectiveness of telemedicine with a reduction of health care costs.

Conclusions: There is promising evidence supporting telemedicine in pediatric settings. However, there is a lack of evaluation of telemedicine in comparison with usual outpatient care for noninferiority of clinical outcomes, and this review highlights the need for a more standardized approach to evaluation of digital interventions.

(J Med Internet Res 2022;24(10):e38267) doi:10.2196/38267

KEYWORDS

telemedicine; telehealth; eHealth; digital health; video consultation; remote consultation; paediatric; child; safeguarding; diabetes; diabetic; asthma; epilepsy; epileptic; renal; kidney; evidence-based medicine; review

Introduction

Telemedicine is the practice of medicine using technology to provide remote health assessment and therapeutic intervention to a patient at a distant site. The spectrum is broad, from simple telephone and video consultations, through wearable digital monitoring, to complex experimental interventions with surgeons guiding robotic instruments to deliver remote surgery.

The adoption of telemedicine consultations escalated rapidly in response to the COVID-19 pandemic [1]. Aside from social distancing, benefits of virtual consultations include potential cost savings and support of sustainability.

Postpandemic, a universal return to face-to-face consultations may risk the loss of some of the advantages of telemedicine. However, rapid implementation and adoption without robust evaluation of usability, efficacy, and effectiveness could potentially lead to suboptimal health outcomes and downstream challenges to providers.

Guidance documents have been published to assist health care professionals to deliver telemedicine [2]. Reviews and evaluations to date have typically focused on specific condition groups and modalities, which does not reflect the variety often encountered in general pediatric outpatient services. Furthermore, there is significant variation in the quality and modality of telemedicine intervention evaluations and potential gaps in outcome measures [3]. A recent broader systematic review of randomized controlled trials (RCTs) in pediatric telemedicine [4] evaluated feasibility, accessibility, satisfaction, and outcomes but did not assess the evaluation of the telemedicine intervention.

Frameworks to benchmark and improve digital health interventions have been developed, for example by the National Institute for Health and Care Excellence (NICE) [5]. These are often designed to evaluate mature interventions and facilitate procurement decisions. In contrast, a World Health Organization (WHO)–published guidance [6] on monitoring and evaluating digital health interventions provides a more fluid framework that can also be applied to novel interventions undergoing iteration as well as more mature interventions being scaled up.

To this end, a group of clinicians with an interest in child health (the Child Health in Practice Group, a voluntary network of UK-based pediatricians) [7] highlighted the need to undertake a high-level scoping review of telemedicine interventions in pediatric outpatient care. The need to understand how children and young people can be effectively supported by emerging technologies was also an outcome finding of the Royal College of Paediatrics and Children Health (RCPCH) 'Paediatrics 2040' project [8]. This review aimed to assess if telemedicine interventions are being evaluated in line with international WHO guidance as well as if there is sufficient evidence to support postpandemic utilization of telemedicine in pediatric settings.

Methods

Study Rationale

The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-SR) statement [9].

At initial study conception, 4 common conditions were examined with the aim to inform clinicians on the suitability of virtual care solutions in outpatient pediatric services. Diabetes was selected for readily measurable biomarkers that can track both long- and short-term disease control (eg, glycated hemoglobin [HbA1c] and blood glucose concentrations), asthma was selected because its biomarkers (eg, peak flow) predominantly reflect short-term disease control, epilepsy is a common condition without a clearly trackable biomarker, and nephrology [10] is an example of a less common patient population whose care is delivered in fewer specialist centers that cover large geographical areas. Initially intended as 4 separate reviews, early in this process, it became apparent that the similarity in methodologies, the number of resulting studies, and the telemedicine modalities in these studies meant that it was more practicable to report findings collectively.

Eligibility Criteria

Studies were selected on the basis of the inclusion and exclusion criteria in Textbox 1.

Textbox 1. Inclusion and exclusion criteria, with rationale, of the studies.

Include:

- Children (0-18 years) receiving outpatient care for any of the following: diabetes, asthma, epilepsy, or renal disease
- Interventions: telemedicine, remote consultation, video call, remote patient monitoring (collecting patient data outside of traditional health care settings to support ongoing patient care)
- Studies comparing interventions with usual outpatient care (including control arms, historic comparisons, or based on user with or without clinician perceptions)
- Studies published between 2011 and the present day
- Studies published in the English language
- Studies that are primary research studies

Exclude:

- Adults (>18 years) or where not possible to disaggregate data for participants younger than 18 years
- Outcome data pertaining to conditions other than diabetes, asthma, epilepsy, or renal disease
- Interventions not intended to replace current outpatient medical health care services:
  - Education
  - Behavioral interventions (eg, cognitive behavioral therapy)
  - Family therapies (where not intended to replace current outpatient health care services)
- Remote patient monitoring not directly related to outpatient care:
  - Support telemetry of electroencephalogram (EEG), electrocardiogram (ECG), or other continuous data from inpatients or transport patients
  - Teleconferencing between health care professionals (eg, tertiary center reviews or multidisciplinary team meetings)
- Studies not published in the English language
- Studies that are conference abstracts, letters, study protocols, systematic reviews, or review articles.

Information Sources and Search Strategy

PubMed, Embase, and CINAHL databases were searched using the search strategies presented in Table S1 in Multimedia Appendix 1 on April 15, 2021, and repeated on May 31, 2022.

Selection Process

The Rayyan [11] web-based tool was used to assist the selection process, initially by identification of duplicates, which required acceptance or rejection by a reviewer. After removal of duplicates, all records were initially screened by title and abstract, performed independently by 2 reviewers assigned to each record. Decisions were unblinded after completion. Where decisions conflicted, this was resolved by discussion or a third reviewer if agreement could not be reached.

Data Collection Process

Data were initially extracted into synthesis tables for each disease group and corroborated by a second reviewer. To ensure good inter-reviewer consistency, a blinded calibration exercise was performed.

Data Items (Outcomes)

For all studies, the following data fields were extracted into the synthesis table:

- Title, author, date of publication, URL, and DOI
- Study design, number of participants, study population, and location of study
- The telemedicine intervention under evaluation and its maturity
- Evidence of impact on pediatric care in 3 domains:
  - Usability and feasibility of telemedicine in pediatric settings
  - Efficacy and effectiveness as evidenced by process and clinical outcomes [12]
  - Implementation science issues
- Confidence in the strength of the evidence in each of these domains

Studies were analyzed for maturity of intervention, risk of bias, and outcomes reported. The WHO guidance [6] was used to determine the maturity of the telemedicine intervention, in turn defining the appropriate focus of evaluation as well as appropriate claims regarding the anticipated benefits of the intervention. Intervention maturity was defined by the size of deployment, intervention setting (controlled/uncontrolled), and what previous testing of the intervention has taken place. Data categories related to the impact on pediatric care were also based on the stages of evaluation outlined in the WHO guidance [6].

Due to the variety of study designs anticipated, formal quality assessment tools (eg, Critical Appraisal Skills Programme or National Institutes of Health National Heart, Lung, and Blood Institute tools) were not used. Instead, the hierarchy of evidence outlined in the WHO guidance [6] was used as a high-level indication of the confidence in the strength of the evaluations'

evidence, categorized as poor, fair, good, or excellent based on the overarching study methodology. This framework was chosen as it enables high-level assessments of evidence across different study designs and meaningful comparison of interventions of different maturities.

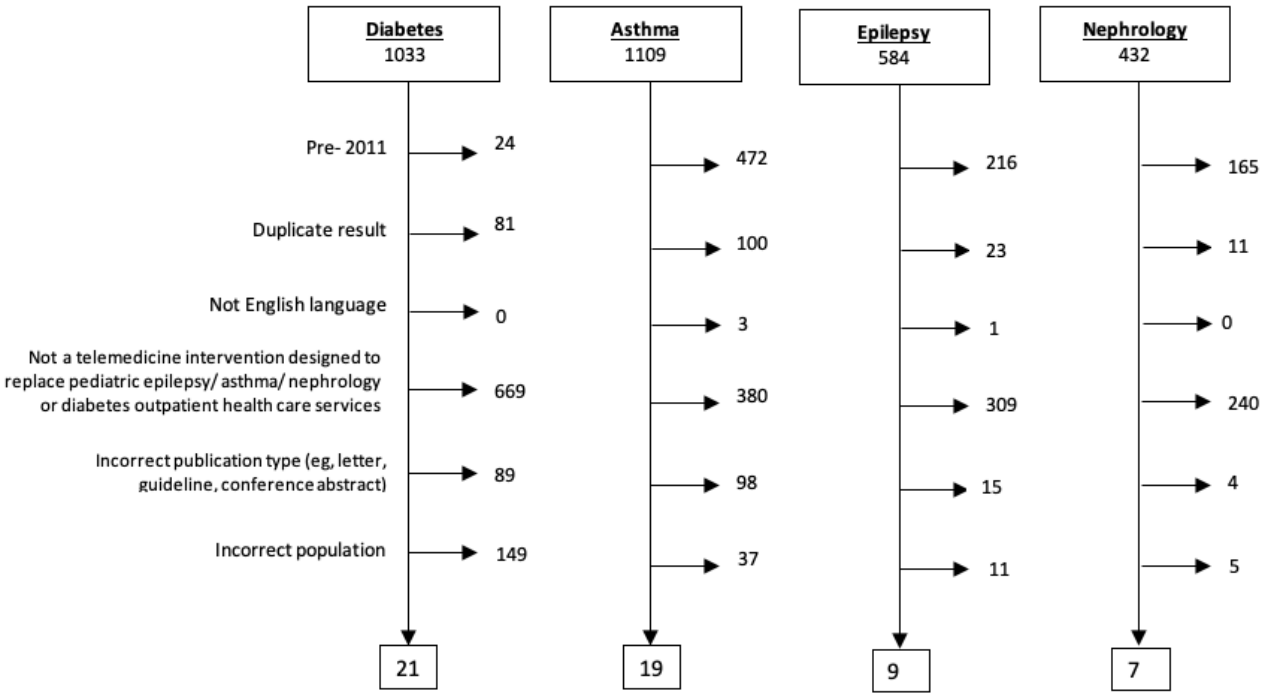
The blinded calibration exercise identified that proportions of agreement for both intervention maturity assessment and strength of evidence were 100% for all extractors.

Study outcome measures and findings were reviewed, collated, and synthesized in tabular form. Studies were examined for reported outcomes (feasibility and usability, clinical and process outcomes, and implementation science issues), and coded as either positive, negative, or equivocal.

Results

Following a priori exclusions (Figure 1), a total of 56 published studies relating to telemedicine in pediatric services were included.

Figure 1. A series of 4 CONSORT diagrams reporting identified articles and reasons for exclusion; a total of 56 articles was suitable for final inclusion and analysis.



Description of Studies

A total of 11 studies were RCTs, of which 3 were multisite, the highest level in the WHO hierarchy of evidence [6]. Other study designs were quasiexperimental studies (9 studies), cohort studies (11 studies), case control studies (4 studies), cross-sectional studies (19 studies), observational studies (1 study), and quality improvement project (1 study).

Of these studies, 49 were quantitative in nature, 3 studies utilized mixed methods, and 4 studies were qualitative.

Interventions and Stage of Maturity

The scale and maturity of the telemedicine deployments varied from prototypes undergoing user testing all the way to large-scale deployments covering multiple sites. Using the WHO 6-stage intervention maturity life cycle [6], the interventions across the primary papers were categorized as prototype or pre-prototype (9 studies), pilot (usually a single deployment in controlled circumstances; 42 studies), demonstration (moderate-scale implementation no longer in controlled settings; 3 studies), and scale-up (intervention that is ready for implementation at subnational or higher level; 2 studies).

A range of telemedicine modalities were utilized, with 46 studies examining a single modality, namely videoconferencing or video calls (22 studies); telephone (4 studies); instant messaging, chatbot, or SMS (3 studies); and remote patient monitoring (RPM; 17 studies). In 6 studies, the RPM platform was hosted on a smartphone application. The remaining 10 studies examined a combination of 2 or more of the modalities.

Main outcomes were feasibility or usability, clinical or process outcomes, and implementation science issues. Figure 2 summarizes the data that were reported and overall findings. Clinical or process outcomes are included only where comparison was made and not if simply reporting the number of events, for example.

Figure 2. Summary of findings for main evaluation domains, classified as "Not reported," "Positive," "Neutral," or "Negative" [13-68].

Disease group	Study	Feasibility and/or usability	Clinical/process outcomes	Implementation science
Asthma	Deschildre et al, 2012[13]	⊕	±	X
	Lin et al, 2014[14]	±	X	X
	Gahleitner et al, 2015[15]	⊕	⊕	X
	Portnoy et al, 2016[16]	⊕	±	X
	van den Wijngaart et al, 2017[17]	X	⊕	X
	Halterman et al, 2018[18]	⊕	⊕	X
	van den Wijngaart et al, 2018[19]	⊕	⊕	⊖
	Bian et al, 2019[20]	X	⊕	X
	Kosse et al, 2019 (Feb)[21]	⊕	⊕	X
	Kosse et al, 2019 (April)[22]	X	±	X
	Lv et al, 2019[23]	X	⊕	⊕
	Mikalsen et al, 2019[24]	⊕	⊕	X
	Bui et al, 2020[25]	⊕	X	X
	Lin et al, 2020[26]	⊕	⊕	X
	Kowatsch et al, 2021[27]	⊕	X	X
	Kumari et al, 2021[28]	⊕	X	X
	MacGeorge et al, 2021[29]	±	X	X
	Makhecha et al, 2021[30]	⊕	±	X
	Mayoral et al, 2021[31]	⊕	X	X
Diabetes	Landau et al, 2012[32]	⊕	±	X
	Choi et al, 2013[33]	X	⊕	X
	Pena et al, 2013[34]	⊕	⊕	X
	Damgaard et al, 2014[35]	⊕	X	X
	Chorianopoulou et al, 2015[36]	⊕	X	X
	Smith et al, 2016[37]	⊕	⊕	X
	Brown et al, 2017[38]	±	⊕	X
	Losiouk et al, 2017[39]	±	X	X
	Gandrud et al, 2018[40]	X	±	X
	Guttmann-Bauman et al, 2018[41]	⊕	±	X
	Burckhardt et al, 2019[42]	⊕	⊕	X
	Döğer et al, 2019[43]	X	⊕	X
	Kaushal et al, 2019[44]	⊕	X	X
	Crossen et al, 2020[45]	⊕	⊕	⊕
	Lim et al, 2020[46]	⊕	X	X
	Odeh et al, 2020[47]	⊕	X	X
	Predieri et al, 2020[48]	X	⊕	X
	von Sengbusch et al, 2020[49]	X	±	X
	Bassi et al, 2022[50]	⊕	X	X
	Frielitz et al, 2022[51]	X	X	⊕
	Troncone et al, 2022[52]	⊕	X	X
Epilepsy	Bahrani et al, 2017[53]	⊕	±	⊕
	Conde-Blanco et al, 2020[54]	⊕	⊕	X
	Dozières-Puyravel et al, 2020[55]	⊕	X	X
	Panda et al, 2020[56]	⊕	X	X
	Nieman et al, 2020[57]	⊕	⊕	⊕
	Trivisano et al, 2020[58]	⊕	±	X
	Anuszkiewicz et al, 2022[59]	±	X	±
	Armeno et al, 2022[60]	⊕	⊕	X
	Gali et al, 2022[61]	⊕	X	⊕
Nephrology	Trnka et al, 2015[62]	X	X	⊕
	Aydemir et al, 2021[63]	⊕	X	X
	Trace et al, 2020[64]	⊕	X	X
	Pooni et al, 2022[65]	⊕	X	X
	Raina et al, 2021[66]	⊕	X	X
	Qiu et al, 2021[67]	⊕	X	X
	Clark et al, 2022[68]	⊕	X	X

Legend:

X Not reported/examined.

⊕ Overall broadly a positive impact.

± Overall broadly a neutral / non-inferior impact.

⊖ Overall broadly a negative impact.

Feasibility and Usability of Telemedicine

Data on feasibility or usability of the telemedicine intervention were reported in 45 studies. All but 1 of the studies not reporting these data looked at a pilot stage intervention, with the remaining study [20] examining a scale-up stage intervention. Technical difficulties were reported in 20 studies, but not all positively or negatively testified to this. A breakdown of the confidence in the strength of the evidence in this evaluation domain across the studies is outlined in Table 1.

Table 1. Summary of confidence in the strength of the evidence of usability or feasibility of telemedicine interventions across each condition group in 45 studies.

Variables	Epilepsy	Nephrology	Diabetes	Asthma	Total
Confidence in the strength of the evidence, number of studies					
Excellent	0	0	0	1	1
Good	3	0	10	11	24
Fair	6	6	5	3	20
Poor	0	0	0	0	0
Number of studies within condition group reporting these data, n (%)	9/9 (100)	6/7 (86)	15/21 (71)	15/19 (79)	45/56 (80)

In the results, there were common positive themes identified from health care professionals, patients, and families in support of the telemedicine interventions, including convenience, perceived cost savings, and ease of use. Interventions were generally well-accepted and would be recommended to others. Use in preference to usual care was explored in 4 studies. In 1 study [63], families reported use of telehealth beyond the coronavirus pandemic was not wanted, and in 3 studies [50,64,68], parents emphasized it was wanted to supplement but not substitute in-person clinics.

There were 16 studies (16/56, 29%) that had requirements for participants to have specific owned mobile devices or internet access or excluded participants for whom technical issues meant consultation could not proceed. Damgaard and Young [35] presented a statistically significant improvement in parental satisfaction following intervention; however, broadband bandwidth was insufficient for many schools with the intervention, necessitating installation of separate internet connections.

For smartphone-based RPM, this evaluation domain was conducted in 4 of 6 studies. In 1 study [21], 96% of pharmacists were satisfied with the intervention, and although 77% of patients felt it was easy to use and 78% would recommend to others, 19% reported technical issues as a reason to not use the intervention. In 3 other studies, patient satisfaction was reported, but health care professionals' satisfaction was not. In 1 pilot study [25], children were satisfied (63%) or very satisfied (32%) with their experience with the app, similarly rated by parents. The interventions by Mikalsen et al [24] had a median score of 18/20 for functionality and overall assessment, while mean System Usability Scale scores in the study by Mayoral et al [31] were 92.9 (0: negative; 100: positive).

The majority of evidence against the use of telemedicine came from studies in asthma services, although this is in the context of otherwise very positive evidence, including from 1 study [13] with excellent confidence in the strength of the evidence. Prototype interventions were examined in 2 studies [14,15], with issues related to the usability of the non-smartphone–based remote monitoring system. A third study [19] looking at scale-up of a non-smartphone–based remote monitoring system found that 4 of 14 sites were unable to successfully implement the intervention, although in only 1 case was a reason (insufficient staff) provided. Finally, a pilot study [28] of telephone consultations for asthma found that only 40% of respondents wished to continue with the modality beyond the pandemic.

Telemedicine Impact on Processes and Clinical Outcomes

Clinical and process outcome data were collected in 31 studies: 23 pilot studies, 4 prototype studies, 2 scale-up studies, and 2 demonstration studies. Of those not reporting outcome data, 19 were pilot studies, 5 were prototype studies, and 1 was a demonstration.

For the studies that collected and reported outcomes data, a breakdown of the confidence in the strength of this evidence is outlined in Table 2.

Outcome data, when provided, were mostly, but not universally, positive. Process measures were more frequently provided than clinical outcomes. Almost all the studies with good or excellent confidence in the strength of the evidence looked at asthma or diabetes.

Table 2. Summary of confidence in the strength of the evidence of clinical or process outcomes of telemedicine interventions across each condition group in 30 studies.

Variables	Epilepsy	Nephrology	Diabetes	Asthma	Total
Confidence in the strength of the evidence, number of studies					
Excellent	0	0	1	2	3
Good	1	0	8	9	18
Fair	4	0	1	0	5
Poor	0	0	2	2	4
Number of studies within condition group reporting these data, n (%)	5/9 (56)	0/7 (0)	12/21 (57)	13/19 (68)	30/56 (54)

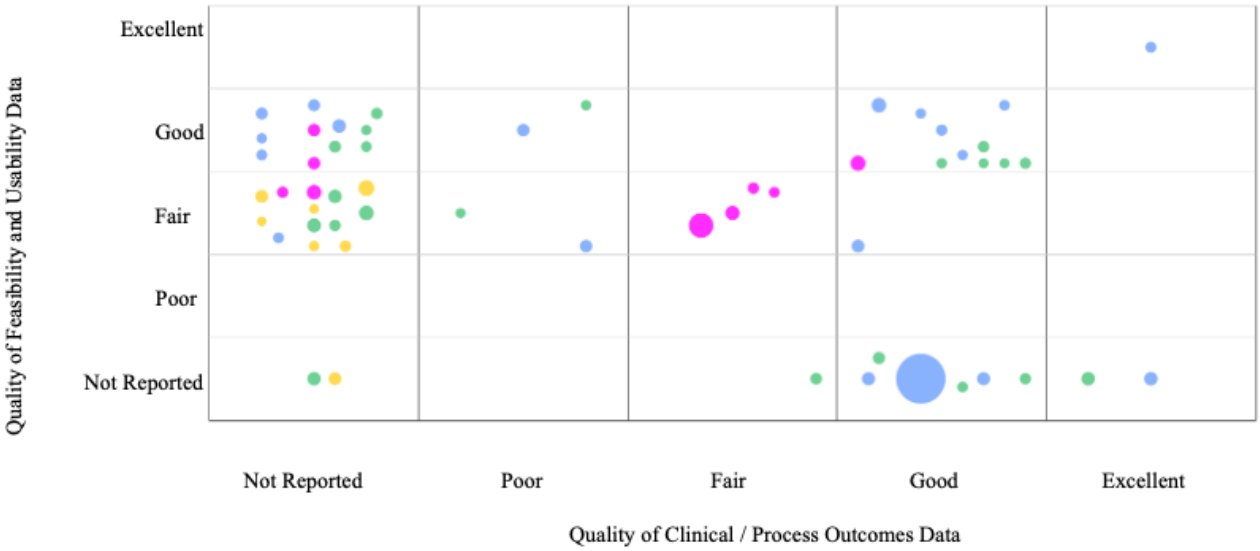
A summary of the outcomes, appraisal, and quality of evidence of each of the 56 studies is presented in Table S2 in Multimedia Appendix 1 [13-68]. Across all subgroups, the telemedicine intervention was acceptable or feasible to the patient, their family, or the health care professionals; however, not all studies reported the view of all parties. Telemedicine interventions were considered beneficial, and for diabetes and asthma particularly, perceived benefits included improved understanding and management of the child's condition, with some studies reporting measures of quality of life. Across all studies, there was limited measurement of standardized clinical outcomes, although these were more commonly reported in asthma (peak flow) and diabetes (HbA1c). Many studies reported interventions adopted as a consequence of the COVID-19 pandemic, and although generally well-accepted in their own right, acceptability in comparison or instead of usual care was not always explored. The study by Gandrud et al [40] was the only study in which clinical outcome data were compared with standard care. Their intervention demonstrated improvement in HbA1c and health-related quality of life in the intervention group, but this was not statistically significant.

One limiting factor common across all groups was exclusions of participants based on lack of access to the internet or an appropriate device.

Confidence in the Strength of the Evidence of Studies

As outlined in Figure 3, 10 studies reported outcome data without examining feasibility or usability, while 23 studies only reported feasibility or usability data. The majority (21/22, 96%) of studies that reported both variables had fair or good confidence in the strength of findings. A solitary study [13] was assessed as having the highest confidence (excellent) for both evidence of feasibility or usability and evidence on outcomes.

Figure 3. Bubble plot depicting the quality of data (per World Health Organization guidance [6]) of studies reporting feasibility and/or usability and outcomes. The size of the bubbles corresponds to the size of the study cohort, and the color indicates the disease category: green: diabetes; blue: asthma; pink: epilepsy; yellow: nephrology.



Implementation Science Issues of Telemedicine

The most common implementation science factor examined was cost-effectiveness, reported in 5 studies (5/56, 9%; details in Table S2 in Multimedia Appendix 1).

One study [53] examining epilepsy services focused on a demonstration of a telephone consultation intervention that identified a saving of 865 INR (approximately US $11.25 per patient). A study in the pilot phase [62] that utilized multiple telemedicine interventions in nephrology patients, identified cost savings, estimated to be US $505 per consultation, mostly associated with reducing travel and accommodation requirements. Finally, a pilot study [23] of a smartphone-based RPM intervention for asthma care found a statistically significant decrease in medical expenses from 1179 RMB to 931 RMB (approximately US $145) per patient. A quasirandomized multicenter study found direct diabetes-associated 6-month costs to be €4702 in the intervention group, compared with €4936 in the control group [51]. Another pilot study [61] found a statistically significant difference (P<.001) in out-of-pocket expenses for telehealth, US $35, compared with US $176 for an in-person visit.

When compared with usual clinics, Gali et al [61] found that, with telehealth, missed school hours were reduced by 49%, missed work hours were reduced by 48%, and mileage was 32 miles compared with 49 miles for in-patient visits (all P<.001). Significant time savings were also identified when using videoconferencing for diabetes [45] and reported (but not quantified) by 98% of respondents using non-smartphone–based remote monitoring for epilepsy [57]. The final implementation science issue identified was in non-smartphone remote monitoring for asthma, for which lack of structural financial reimbursement was identified as the main barrier [19].

Discussion

Principal Findings

This review examined the existing literature regarding pediatric telemedicine interventions in asthma, diabetes, epilepsy, and nephrology. These conditions cover a mixture of common and

rare diseases, with and without availability of biomarkers including short- and long-term control. Reported clinical outcomes were heterogeneous, making pooling of results impossible. A wide variety of outcome and process measures was used, with no clear standardization, even within condition groups. This also made cost-effectiveness analysis challenging. Although several studies did report cost-effectiveness, this was usually heavily caveated and thus not conducive to supporting larger commissioning or service redesign decisions.

With regards to how digital interventions are used and perceived, the majority of identified studies examined the evidence on usability and feasibility of the telemedicine intervention. Satisfaction levels among clinicians and patients were high, with users open to and often enthusiastic for future use of telemedicine. The confidence in the strength of these findings was most typically fair or good. In no study were there observed disagreements between the sentiment of professionals and those of patients and families. Further research should build on this evidence base, and we recommend that research and evaluation frameworks encourage the standardized collection of usability and acceptability data, particularly in prototype, pilot, and demonstration stage interventions.

Where problems with the usability and feasibility of telemedicine were identified, this was primarily due to technical issues with video conferencing as well as non-smartphone–based RPM. In one instance [19], wider implementation issues (eg, lack of staff) prevented successful scale-up of an asthma non-smartphone–based RPM intervention in 4 of 14 sites and identifies lack of structural financial reimbursement of web-based monitoring as a significant obstacle in diffusion of eHealth innovation. Such challenges of large-scale digital implementation are well known, with national reports suggesting that the process can take several years to fully iterate [69]. In contrast, smartphone-based RPM had particularly positive evidence, with clinician and patient satisfaction levels over 90% and good confidence in the strength of the evidence in one-half of these studies.

With regards to the impact of digital interventions, the majority of literature found that outcomes improved or were equivocal to traditional care. Where there was evidence of negative impact on clinical outcomes, this was not statistically significant. The strength in the confidence in clinical outcome evidence varied across the literature. There was poor evidence in nephrology. In contrast, in diabetes and asthma, conditions with established biomarkers that can objectively monitor disease progression, there were several studies with excellent confidence in the evidence. Unfortunately, this was not always coupled with good quality evidence on the perception and use of the intervention. Only 18 studies reported good or excellent confidence in the evidence for both how the studied intervention was received and its impact.

The potential long-term benefits of telemedicine include decreased travel (time and expense) for families, reduced exposure to nosocomial infection for vulnerable patients, and reductions in the carbon footprint of health care [70]. The COVID-19 pandemic has seen the rapid implementation of telemedicine interventions with many of these benefits realized over the past 2 years [1]. Improvements in digital implementation could represent a boon for its wider adoption, but caution should prompt the collection of meaningful data to ensure these are noninferior to traditional modalities of care.

Among the potential limitations are a lack of physical examination and the resulting impact on clinical decision-making, privacy concerns, and the impact of digital exclusion [63]. However, reassuring outcome data and user satisfaction, in line with WHO guidance [6], can provide assurance in these regards. One important reflection is the lack of child protection or safeguarding literature in the area [3].

Evidence From Previous Literature

The search protocols identified 5 previous reviews [71-75] focused on telemedicine in diabetes and not included in the final synthesis. These papers examined a variety of modalities including video consultations, telephone consultations, text services, and RPM, all in the pilot stage. Of these reviews, 4 [71,73-75] identified data on feasibility and usability, with findings consistent with this synthesis. With the exception of a small number of technical problems with 1 intervention's GPRS wireless system [71], findings were universally positive. Identified benefits included improved access to care, increased parental satisfaction, and perceived time savings.

Four reviews identified outcome data [71-74], 2 of which reported statistically significant improvements in HbA1c and emergency department visits [72,74]. Two reviews noted that telephone, SMS services, and non-smartphone–based RPM [71,73] had no impact on HbA1c. One of these reviews [71] identified that telephone and SMS services may improve patient engagement and self-efficacy.

Recommendations

The review identifies an urgent need for a more standardized approach to evaluation of digital interventions. There is a lack of literature examining this area despite the increasing adoption of such virtual consultations. Much of the literature does not include meaningful data on usability and feasibility of the intervention recommended by the WHO guidance [6] and particularly important for early, pilot-stage interventions.

Although usability is an important measure, it is also important to evaluate changes in health care practices for noninferiority for clinical outcomes. Evaluations should be designed to review meaningful clinical outcomes, for example, in pediatric nephrology, the rate of progression of renal impairment and transplant survival. Proxy measures for these could also include proteinuria and medication concordance. Standardization, perhaps through an agreed outcome set, would enable interventions to be compared and results pooled. Professional organizations such as the RCPCH can lead this to produce evaluation frameworks, facilitating scientific rigor among suppliers to undertake high-quality evaluations.

For clinicians interested in digital implementation, the early findings across these studies are promising, particularly in smartphone-based RPM and video consultation. Clinicians who are implementing or piloting digital interventions should focus on building robust evaluation strategies, in line with established

guidance. Additions to the evidence base should focus on promoting higher-quality studies, ideally RCT or other experimental study designs.

High-quality evaluation can be promoted through restructuring innovation funds, which should reward comprehensive evaluation strategies aligned with international guidance and which should ideally set aside a portion of funding to be used exclusively for evaluation.

Limitations

Although the review is a starting point for further evaluation and research, a number of limitations should be acknowledged. The focus on 4 disease groups, a pragmatic amalgamation of initially parallel reviews, provides breadth but is not complete. The review is also at high risk of publication bias. Although we found several examples of prototype interventions with mixed results, other interventions that may have had unfavorable results may have been excluded from publication. A common theme in studies was the potential exclusion of participants who did not have access to appropriate technology, and this may have implications of health inequalities in this space.

The WHO framework [6] offers a high-level assessment of digital health interventions, enabling comparison between different study designs. However, it does not differentiate between the quality of similar designs. For example, a poorly designed multisite RCT may offer less compelling evidence than a well-designed longitudinal study, which would not be reflected in our chosen approach. Nonetheless, the framework is well-suited to assess interventions of varying maturity and enables some comparison between different study designs.

Conclusion

Current evidence indicates that, across a range of modalities, including telephone or video calls, text messaging, and more comprehensive RPM systems, telemedicine is viewed as an acceptable tool to deliver pediatric outpatient care. Although promising, existing results should be taken with consideration of the data's limitations. When telemedicine interventions are to be implemented, appropriate gathering of data is needed to secure an evidence base that interventions are safe and not associated with inferior clinical outcomes. Outcome measures should include child safety and clinical outcomes to ensure noninferiority to traditional face-to-face consultation.

Acknowledgments

We thank Dr Rebecca Hewitson and Dr Saranya Ravindran for help in screening some of the studies for inclusion.

Authors' Contributions

All authors contributed to the study design and conceptualization, had full access to all the data in the study, reviewed and edited the final manuscript draft, and had final responsibility for the decision to submit for publication. GS performed the literature search. GS, MJH, AAY, and HL screened the results and extracted the data. GS and AAY accessed the data, and GS, AAY, DR, and KP verified the data. GS, AAY, and MJH analyzed and interpreted the data and wrote the original manuscript draft.

Conflicts of Interest

GS, MJH, HL, DR, and KP declare no competing interests. AAY is a paid employee (part-time) at Huma Therapeutics (a digital health company) since September 2020.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 72 KB - jmir_v24i10e38267_app1.docx]

References

1. Friedman AB, Gervasi S, Song H, Bond AM, Chen AT, Bergman A, et al. Telemedicine catches on: changes in the utilization of telemedicine services during the COVID-19 pandemic. Am J Manag Care 2022 Jan 01;28(1):e1-e6 [FREE Full text] [doi: 10.37765/ajmc.2022.88771] [Medline: 35049260]
2. Principles for conducting virtual consultations with children and young people. Royal College of Paediatrics and Child Health. 2021 Jan 05. URL: https://www.rcpch.ac.uk/resources/principles-conducting-virtual-consultations-children-young-people#during-the-virtual-consultation [accessed 2022-10-18]
3. Southgate G, Harmer MJ, Hewitson R, Livesey H, Longbottom K, Yassaee A, et al. Safeguarding should be a key outcome measure when evaluating video consultations in paediatrics. J Paediatr Child Health 2021 May;57(5):766-767. [doi: 10.1111/jpc.15494] [Medline: 33856068]
4. Shah AC, Badawy SM. Telemedicine in pediatrics: systematic review of randomized controlled trials. JMIR Pediatr Parent 2021 Feb 24;4(1):e22696 [FREE Full text] [doi: 10.2196/22696] [Medline: 33556030]
5. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence (NICE). 2018 Dec 10. URL: https://www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-pdf-1124017457605 [accessed 2022-10-18]

6. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252183/9789241511766-eng.pdf?sequence=1 [accessed 2021-09-09]
7. Pye S. The Running Horse Group: inspiring the child health workforce. Q. 2018 Jul 20. URL: https://q.health.org.uk/blog-post/the-running-horse-group-welcoming-a-child-health-sig-within-q/ [accessed 2022-02-03]
8. Summary of our work. Paediatrics 2040. URL: https://paediatrics2040.rcpch.ac.uk/summary-of-our-work/ [accessed 2022-03-03]
9. 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]
10. Improving the standard of care of children with kidney disease through paediatric nephrology networks. Royal College of Paediatrics and Child Health. 2011 Aug. URL: https://www.rcpch.ac.uk/sites/default/files/Guidance_on_paediatric_nephrology_networks.pdf [accessed 2022-10-18]
11. 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]
12. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
13. Deschildre A, Béghin L, Salleron J, Iliescu C, Thumerelle C, Santos C, et al. Home telemonitoring (forced expiratory volume in 1 s) in children with severe asthma does not reduce exacerbations. Eur Respir J 2012 Feb;39(2):290-296 [FREE Full text] [doi: 10.1183/09031936.00185310] [Medline: 21852334]
14. Lin H, Chiang L, Wen T, Yeh K, Huang J. Development of online diary and self-management system on e-Healthcare for asthmatic children in Taiwan. Comput Methods Programs Biomed 2014 Oct;116(3):299-310. [doi: 10.1016/j.cmpb.2014.05.004] [Medline: 24947614]
15. Gahleitner F, Legg J, Holland E, Pearson S, Roberts G. The validity and acceptability of a text-based monitoring system for pediatric asthma studies. Pediatr Pulmonol 2016 Jan;51(1):5-12. [doi: 10.1002/ppul.23200] [Medline: 25847724]
16. Portnoy JM, Waller M, De Lurgio S, Dinakar C. Telemedicine is as effective as in-person visits for patients with asthma. Ann Allergy Asthma Immunol 2016 Sep;117(3):241-245. [doi: 10.1016/j.anai.2016.07.012] [Medline: 27613456]
17. van den Wijngaart LS, Roukema J, Boehmer ALM, Brouwer ML, Hugen CAC, Niers LEM, et al. A virtual asthma clinic for children: fewer routine outpatient visits, same asthma control. Eur Respir J 2017 Oct;50(4):1 [FREE Full text] [doi: 10.1183/13993003.00471-2017] [Medline: 28982775]
18. Halterman JS, Fagnano M, Tajon RS, Tremblay P, Wang H, Butz A, et al. Effect of the school-based telemedicine enhanced asthma management (SB-TEAM) program on asthma morbidity: a randomized clinical trial. JAMA Pediatr 2018 Mar 05;172(3):e174938 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.4938] [Medline: 29309483]
19. van den Wijngaart LS, Geense WW, Boehmer AL, Brouwer ML, Hugen CA, van Ewijk BE, et al. Barriers and facilitators when implementing web-based disease monitoring and management as a substitution for regular outpatient care in pediatric asthma: qualitative survey study. J Med Internet Res 2018 Oct 30;20(10):e284 [FREE Full text] [doi: 10.2196/jmir.9245] [Medline: 30377147]
20. Bian J, Cristaldi KK, Summer AP, Su Z, Marsden J, Mauldin PD, et al. Association of a school-based, asthma-focused telehealth program with emergency department visits among children enrolled in South Carolina Medicaid. JAMA Pediatr 2019 Nov 01;173(11):1041-1048 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.3073] [Medline: 31498379]
21. Kosse RC, Bouvy ML, de Vries TW, Koster ES. Effect of a mHealth intervention on adherence in adolescents with asthma: A randomized controlled trial. Respir Med 2019 Mar;149:45-51 [FREE Full text] [doi: 10.1016/j.rmed.2019.02.009] [Medline: 30803885]
22. Kosse RC, Bouvy ML, de Vries TW, Koster ES. Evaluation of a mobile health intervention to support asthma self-management and adherence in the pharmacy. Int J Clin Pharm 2019 Apr;41(2):452-459 [FREE Full text] [doi: 10.1007/s11096-019-00798-3] [Medline: 31028598]
23. Lv S, Ye X, Wang Z, Xia W, Qi Y, Wang W, et al. A randomized controlled trial of a mobile application-assisted nurse-led model used to improve treatment outcomes in children with asthma. J Adv Nurs 2019 Nov;75(11):3058-3067. [doi: 10.1111/jan.14143] [Medline: 31241192]
24. Mikalsen IB, Nassehi D, Øymar K. Vortex whistle and smart phone application for peak flow recordings in asthmatic children: a feasibility study. Telemed J E Health 2019 Nov;25(11):1077-1082 [FREE Full text] [doi: 10.1089/tmj.2018.0270] [Medline: 30570372]
25. Bui AAT, Hosseini A, Rocchio R, Jacobs N, Ross MK, Okelo S, et al. Biomedical REAl-Time Health Evaluation (BREATHE): toward an mHealth informatics platform. JAMIA Open 2020 Jul;3(2):190-200 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa011] [Medline: 32734159]
26. Lin NY, Ramsey RR, Miller JL, McDowell KM, Zhang N, Hommel K, et al. Telehealth delivery of adherence and medication management system improves outcomes in inner-city children with asthma. Pediatr Pulmonol 2020 Apr;55(4):858-865 [FREE Full text] [doi: 10.1002/ppul.24623] [Medline: 31905264]

27. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
28. Kumari J, Jat KR, Kabra SK. Role of telemedicine in follow-up care of children with respiratory illnesses at a tertiary care hospital - an ambispective observational study. Indian J Pediatr 2021 Oct;88(10):974-978 [FREE Full text] [doi: 10.1007/s12098-020-03590-8] [Medline: 33394296]
29. MacGeorge CA, King K, Andrews AL, Sterba K, Johnson E, Brinton DL, et al. School nurse perception of asthma care in school-based telehealth. J Asthma 2022 Jun;59(6):1248-1255. [doi: 10.1080/02770903.2021.1904978] [Medline: 33730979]
30. Makhecha S, Jamalzadeh A, Irving S, Hall P, Sonnappa S, Saglani S, et al. Paediatric severe asthma biologics service: from hospital to home. Arch Dis Child 2021 Sep;106(9):900-902 [FREE Full text] [doi: 10.1136/archdischild-2020-320626] [Medline: 33589471]
31. Mayoral K, Garin O, Caballero-Rabasco MA, Praena-Crespo M, Bercedo A, Hernandez G, ARCA group. Smartphone app for monitoring asthma in children and adolescents. Qual Life Res 2021 Nov;30(11):3127-3144. [doi: 10.1007/s11136-020-02706-z] [Medline: 33387290]
32. Landau Z, Mazor-Aronovitch K, Boaz M, Blaychfeld-Magnazi M, Graph-Barel C, Levek-Motola N, et al. The effectiveness of Internet-based blood glucose monitoring system on improving diabetes control in adolescents with type 1 diabetes. Pediatr Diabetes 2012 Mar;13(2):203-207. [doi: 10.1111/j.1399-5448.2011.00800.x] [Medline: 21848925]
33. Choi YS, Berry-Cabán C, Nance J. Telemedicine in paediatric patients with poorly controlled type 1 diabetes. J Telemed Telecare 2013 Jun;19(4):219-221. [doi: 10.1258/jtt.2012.120506] [Medline: 23401606]
34. Peña NV, Torres M, Cardona JAC, Iniesta R. Impact of telemedicine assessment on glycemic variability in children with type 1 diabetes mellitus. Diabetes Technol Ther 2013 Feb;15(2):136-142. [doi: 10.1089/dia.2012.0243] [Medline: 23289433]
35. Damgaard G, Young L. Virtual nursing care for school children with diabetes. Journal of Nursing Regulation 2014 Jan;4(4):15-24. [doi: 10.1016/s2155-8256(15)30106-x]
36. Chorianopoulou A, Lialiou P, Mechili E, Mantas J, Diomidous M. Investigation of the quality and effectiveness of telemedicine in children with diabetes. Stud Health Technol Inform 2015;210:105-109. [Medline: 25991111]
37. Smith NM, Satyshur RD. Pediatric diabetes telemedicine program improves access to care for rural families: role of APRNs. Pediatr Nurs 2016;42(6):294-299. [Medline: 29406653]
38. Brown C, Mulhearn H. The Type 1 Kidz project. Part Two: Introducing a simple telehealth system. Journal of Diabetes Nursing 2017 Jan 30;21(1):32 [FREE Full text]
39. Losiouk E, Lanzola G, Del Favero S, Boscari F, Messori M, Rabbone I, et al. Parental evaluation of a telemonitoring service for children with Type 1 Diabetes. J Telemed Telecare 2018 Apr;24(3):230-237. [doi: 10.1177/1357633X17695172] [Medline: 28345384]
40. Gandrud L, Altan A, Buzinec P, Hemphill J, Chatterton J, Kelley T, et al. Intensive remote monitoring versus conventional care in type 1 diabetes: A randomized controlled trial. Pediatr Diabetes 2018 Feb 21:1. [doi: 10.1111/pedi.12654] [Medline: 29464831]
41. Guttmann-Bauman I, Kono J, Lin AL, Ramsey KL, Boston BA. Use of telehealth videoconferencing in pediatric type 1 diabetes in Oregon. Telemed J E Health 2018 Jan;24(1):86-88. [doi: 10.1089/tmj.2017.0072] [Medline: 28654350]
42. Burckhardt M, Fried L, Bebbington K, Hancock M, Nicholas JA, Roberts A, et al. Use of remote monitoring with continuous glucose monitoring in young children with Type 1 diabetes: the parents' perspective. Diabet Med 2019 Nov;36(11):1453-1459. [doi: 10.1111/dme.14061] [Medline: 31257642]
43. Döğer E, Bozbulut R, Soysal Acar A, Ercan, Kılınç Uğurlu A, Akbaş ED, et al. Effect of telehealth system on glycemic control in children and adolescents with type 1 diabetes. J Clin Res Pediatr Endocrinol 2019 Feb 20;11(1):70-75 [FREE Full text] [doi: 10.4274/jcrpe.galenos.2018.2018.0017] [Medline: 30015620]
44. Kaushal T, Montgomery KA, Simon R, Lord K, Dougherty J, Katz LEL, et al. MyDiaText™: feasibililty and functionality of a text messaging system for youth with type 1 diabetes. Diabetes Educ 2019 Jun 22;45(3):253-259 [FREE Full text] [doi: 10.1177/0145721719837895] [Medline: 30902038]
45. Crossen S, Glaser N, Sauers-Ford H, Chen S, Tran V, Marcin J. Home-based video visits for pediatric patients with poorly controlled type 1 diabetes. J Telemed Telecare 2020 Jul;26(6):349-355 [FREE Full text] [doi: 10.1177/1357633X19828173] [Medline: 30871408]
46. Lim ST, Yap F, Chin X. Bridging the needs of adolescent diabetes care during COVID-19: a nurse-led telehealth initiative. J Adolesc Health 2020 Oct;67(4):615-617 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.012] [Medline: 32798101]
47. Odeh R, Gharaibeh L, Daher A, Kussad S, Alassaf A. Caring for a child with type 1 diabetes during COVID-19 lockdown in a developing country: Challenges and parents' perspectives on the use of telemedicine. Diabetes Res Clin Pract 2020 Oct;168:108393 [FREE Full text] [doi: 10.1016/j.diabres.2020.108393] [Medline: 32858098]
48. Predieri B, Leo F, Candia F, Lucaccioni L, Madeo SF, Pugliese M, et al. Glycemic control improvement in Italian children and adolescents with type 1 diabetes followed through telemedicine curing lockdown due to the COVID-19 pandemic. Front Endocrinol (Lausanne) 2020;11:595735 [FREE Full text] [doi: 10.3389/fendo.2020.595735] [Medline: 33424771]
49. von Sengbusch S, Eisemann N, Mueller-Godeffroy E, Lange K, Doerdelmann J, Erdem A, et al. Outcomes of monthly video consultations as an add-on to regular care for children with type 1 diabetes: A 6-month quasi-randomized clinical

trial followed by an extension phase. Pediatr Diabetes 2020 Dec;21(8):1502-1515. [doi: 10.1111/pedi.13133] [Medline: 33009690]
50. Bassi M, Strati MF, Parodi S, Lightwood S, Rebora C, Rizza F, et al. Patient satisfaction of telemedicine in pediatric and young adult type 1 diabetes patients during Covid-19 pandemic. Front Public Health 2022 Mar 22;10:857561 [FREE Full text] [doi: 10.3389/fpubh.2022.857561] [Medline: 35392480]
51. Frielitz FS, Eisemann N, Werner K, Hiort O, Katalinic A, Lange K, et al. Direct costs of healthcare for children with type 1 diabetes using a CGM system: a health economic analysis of the VIDIKI telemedicine study in a German setting. Exp Clin Endocrinol Diabetes 2022 Sep 03;130(9):614-620. [doi: 10.1055/a-1708-3134] [Medline: 34979571]
52. Troncone A, Cascella C, Chianese A, Zanfardino A, Casaburo F, Piscopo A, et al. Doctor-patient relationship in synchronous/real-time video-consultations and in-person visits: an investigation of the perceptions of young people with type 1 diabetes and their parents during the COVID-19 pandemic. Int J Behav Med 2022 Oct 25;29(5):638-647 [FREE Full text] [doi: 10.1007/s12529-021-10047-5] [Medline: 35076861]
53. Bahrani K, Singh MB, Bhatia R, Prasad K, Vibha D, Shukla G, et al. Telephonic review for outpatients with epilepsy-A prospective randomized, parallel group study. Seizure 2017 Dec;53:55-61 [FREE Full text] [doi: 10.1016/j.seizure.2017.11.003] [Medline: 29127858]
54. Conde-Blanco E, Centeno M, Tio E, Muriana D, García-Peñas JJ, Serrano P, et al. Emergency implementation of telemedicine for epilepsy in Spain: Results of a survey during SARS-CoV-2 pandemic. Epilepsy Behav 2020 Oct;111:107211 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107211] [Medline: 32540769]
55. Dozières-Puyravel B, Auvin S. Usefulness, limitations, and parental opinion about teleconsultation for rare pediatric epilepsies. Epilepsy Behav 2021 Feb;115:107656. [doi: 10.1016/j.yebeh.2020.107656] [Medline: 33317938]
56. Panda PK, Dawman L, Panda P, Sharawat IK. Feasibility and effectiveness of teleconsultation in children with epilepsy amidst the ongoing COVID-19 pandemic in a resource-limited country. Seizure 2020 Oct;81:29-35 [FREE Full text] [doi: 10.1016/j.seizure.2020.07.013] [Medline: 32712376]
57. Niemann MH, Alvarado MC, Camayd-Muñoz C, Jonas RR, Wenger JK, Douglass LM. A multimodal telehealth strategy to improve pediatric epilepsy care. Pediatr Clin North Am 2020 Aug;67(4):629-634. [doi: 10.1016/j.pcl.2020.04.004] [Medline: 32650859]
58. Trivisano M, Specchio N, Pietrafusa N, Calabrese C, Ferretti A, Ricci R, et al. Impact of COVID-19 pandemic on pediatric patients with epilepsy - The caregiver perspective. Epilepsy Behav 2020 Dec;113:107527 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107527] [Medline: 33242768]
59. Anuszkiewicz K, Stogowski P, Zawadzka M, Waszak P, Sokolewicz E, Dułak NA, et al. COVID-19 pandemic influence on epilepsy course in pediatric patients. Epilepsy Behav 2022 Apr;129:108581 [FREE Full text] [doi: 10.1016/j.yebeh.2022.108581] [Medline: 35203013]
60. Armeno M, Caballero E, Verini A, Reyes G, Galarza N, Cresta A, et al. Telemedicine- versus outpatient-based initiation and management of ketogenic diet therapy in children with drug-resistant epilepsy during the COVID-19 pandemic. Seizure 2022 May;98:37-43. [doi: 10.1016/j.seizure.2022.03.023] [Medline: 35417828]
61. Gali K, Joshi S, Hueneke S, Katzenbach A, Radecki L, Calabrese T, et al. Barriers, access and management of paediatric epilepsy with telehealth. J Telemed Telecare 2022 Apr;28(3):213-223 [FREE Full text] [doi: 10.1177/1357633X20969531] [Medline: 33183129]
62. Trnka P, White MM, Renton WD, McTaggart SJ, Burke JR, Smith AC. A retrospective review of telehealth services for children referred to a paediatric nephrologist. BMC Nephrol 2015 Aug 01;16:125 [FREE Full text] [doi: 10.1186/s12882-015-0127-0] [Medline: 26231174]
63. Aydemir S, Ocak S, Saygılı S, Hopurcuoğlu D, Haşlak F, Kıykım E, et al. Telemedicine applications in a tertiary pediatric hospital in Turkey during COVID-19 pandemic. Telemed J E Health 2021 Oct;27(10):1180-1187. [doi: 10.1089/tmj.2020.0381] [Medline: 33301373]
64. Trace SL, Collinson A, Searle AJ, Lithander FE. Using videoconsultations to deliver dietary advice to children with chronic kidney disease: a qualitative study of parent and child perspectives. J Hum Nutr Diet 2020 Dec;33(6):881-889. [doi: 10.1111/jhn.12750] [Medline: 32239728]
65. Pooni R, Pageler NM, Sandborg C, Lee T. Pediatric subspecialty telemedicine use from the patient and provider perspective. Pediatr Res 2022 Jan;91(1):241-246 [FREE Full text] [doi: 10.1038/s41390-021-01443-4] [Medline: 33753896]
66. Raina R, Nair N, Yap HK, Filler G, Sethi SK, Bagga A, et al. Survey of telemedicine by pediatric nephrologists during the COVID-19 pandemic. Kidney Int Rep 2021 Sep;6(9):2316-2322. [doi: 10.1016/j.ekir.2021.06.026] [Medline: 34514192]
67. Qiu Y, Coulson S, McIntyre CW, Wile B, Filler G. Adolescent and caregiver attitudes towards telemedicine use in pediatric nephrology. BMC Health Serv Res 2021 Jun 01;21(1):537 [FREE Full text] [doi: 10.1186/s12913-021-06506-0] [Medline: 34074281]
68. Clark SL, Begin B, De Souza HG, Mallett K, Hanna MG, Richardson T, et al. Telehealth survey of providers and caregivers of children on peritoneal dialysis during the COVID-19 pandemic. Pediatr Nephrol 2022 Apr 14:1 [FREE Full text] [doi: 10.1007/s00467-022-05543-z] [Medline: 35425999]

69. Wachter RM. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England. National Advisory Group on Health Information Technology. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2022-10-18]
70. Holmner A, Ebi KL, Lazuardi L, Nilsson M. Carbon footprint of telemedicine solutions--unexplored opportunity for reducing carbon emissions in the health sector. PLoS One 2014;9(9):e105040 [FREE Full text] [doi: 10.1371/journal.pone.0105040] [Medline: 25188322]
71. Sheehy S, Cohen G, Owen KR. Self-management of diabetes in children and young adults using technology and smartphone applications. Curr Diabetes Rev 2014;10(5):298-301. [doi: 10.2174/1573399810666141010113050] [Medline: 25311195]
72. Sheikhtaheri A, Kermani F. Telemedicine in diagnosis, treatment and management of diseases in children. Stud Health Technol Inform 2018;248:148-155. [Medline: 29726431]
73. Viana LV, Gomes MB, Zajdenverg L, Pavin EJ, Azevedo MJ, Brazilian Type 1 Diabetes Study Group. Interventions to improve patients' compliance with therapies aimed at lowering glycated hemoglobin (HbA1c) in type 1 diabetes: systematic review and meta-analyses of randomized controlled clinical trials of psychological, telecare, and educational interventions. Trials 2016 Feb 17;17:94 [FREE Full text] [doi: 10.1186/s13063-016-1207-6] [Medline: 26888087]
74. Reynolds CA, Maughan ED. Telehealth in the school setting: an integrative review. J Sch Nurs 2015 Feb;31(1):44-53. [doi: 10.1177/1059840514540534] [Medline: 24970886]
75. De Guzman KR, Snoswell CL, Taylor ML, Senanayake B, Haydon HM, Batch JA, et al. A systematic review of pediatric telediabetes service models. Diabetes Technol Ther 2020 Aug;22(8):623-638. [doi: 10.1089/dia.2019.0489] [Medline: 32027176]

Abbreviations

HbA1c: glycated hemoglobin
NICE: National Institute for Health and Care Excellence
PRISMA-SR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCPCH: Royal College of Paediatrics and Child Health
RCT: randomized controlled trial
RPM: remote patient monitoring
WHO: World Health Organization

Edited by R Kukafka; submitted 25.03.22; peer-reviewed by A Sheikhtaheri, R Ramsey; comments to author 28.05.22; revised version received 21.07.22; accepted 02.09.22; published 26.10.22.

Please cite as:
Southgate G, Yassaee AA, Harmer MJ, Livesey H, Pryde K, Roland D
Use of Telemedicine in Pediatric Services for 4 Representative Clinical Conditions: Scoping Review
J Med Internet Res 2022;24(10):e38267
URL: https://www.jmir.org/2022/10/e38267
doi:10.2196/38267
PMID:36287609

Review

Effectiveness of Computerized Cognitive Training in Delaying Cognitive Function Decline in People With Mild Cognitive Impairment: Systematic Review and Meta-analysis

Ran Li[1,2*], MSc, PhD; Jiawei Geng[2*], PhD; Runze Yang[1], MSc; Yumeng Ge[2], BA; Therese Hesketh[1,2], Prof Dr

[1]Institute for Global Health, University College London, London, United Kingdom

[2]Centre for Global Health, School of Public Health, Zhejiang University, Hangzhou Zhejiang, China

[*]these authors contributed equally

Corresponding Author:
Therese Hesketh, Prof Dr
Institute for Global Health
University College London
30 Guilford Street
London, WC1N 1EH
United Kingdom
Phone: 44 20 7242 9789
Email: t.hesketh@ucl.ac.uk

Abstract

Background: With no current cure for mild cognitive impairment (MCI), delaying its progression could significantly reduce the disease burden and improve the quality of life for patients with MCI. Computerized cognitive training (CCT) has recently become a potential instrument for improvement of cognition. However, the evidence for its effectiveness remains limited.

Objective: This systematic review aims to (1) analyze the efficacy of CCT on cognitive impairment or cognitive decline in patients with MCI and (2) analyze the relationship between the characteristics of CCT interventions and cognition-related health outcomes.

Methods: A systematic search was performed using MEDLINE, Cochrane, Embase, Web of Science, and Google Scholar. Full texts of randomized controlled trials of CCT interventions in adults with MCI and published in English language journals between 2010 and 2021 were included. Overall global cognitive function and domain-specific cognition were pooled using a random-effects model. Sensitivity analyses were performed to determine the reasons for heterogeneity and to test the robustness of the results. Subgroup analyses were performed to identify the relationship between the characteristics of CCT interventions and cognition-related effectiveness.

Results: A total of 18 studies with 1059 participants were included in this review. According to the meta-analysis, CCT intervention provided a significant but small increase in global cognitive function compared to that in the global cognitive function of the control groups (standardized mean difference=0.54, 95% CI 0.35-0.73; I^2=38%). CCT intervention also resulted in a marginal improvement in domain-specific cognition compared to that in the control groups, with moderate heterogeneity. Subgroup analyses showed consistent improvement in global cognitive behavior in the CCT intervention groups.

Conclusions: This systematic review suggests that CCT interventions could improve global cognitive function in patients with MCI. Considering the relatively small sample size and the short treatment duration in all the included studies, more comprehensive trials are needed to quantify both the impact of CCT on cognitive decline, especially in the longer term, and to establish whether CCT should be recommended for use in clinical practice.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021278884; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=278884

(J Med Internet Res 2022;24(10):e38624) doi:10.2196/38624

KEYWORDS

computerized cognitive training; mild cognitive impairment

Introduction

Mild Cognitive Impairment

The older adult population is increasing worldwide. In 2017, 962 million people or 13% of the global population were over 60 years of age, and this figure is predicted to rise to 1.4 billion by 2030 [1]. This raises concerns about the growing global burden of degenerative disorders, especially dementia. The development of interventions to prevent, delay, and treat dementia is now recognized as a matter of urgency [2]. Mild cognitive impairment (MCI) is recognized as an intermediary phase between the cognitive changes of normal aging and the onset of dementia, suggesting that it may represent an opportune time to prevent or delay the onset of dementia. Petersen et al [3] reported that globally, an estimated 14.9% of people with MCI aged over 60 years progressed to dementia in the following 2 years and one-third of people living with MCI develop dementia within 5 years [4]. The prevalence of MCI is estimated to be around 16% in adults aged over 60 years, with the risk increasing with age [2]. The diagnostic criteria for MCI include a change in cognition, abnormal cognitive function in one or more domains, but without notable interference in everyday functioning [5].

Currently, there is no specific diagnostic test for MCI. However, global cognitive function is measured most commonly using the Mini-Mental State Examination and the Montreal Cognitive Assessment [6]. Measures such as executive function, working memory, episodic memory, and quality of life are also commonly used. In this context, executive function is the most complex cognitive process necessary for goal-directed behavior [7]. Working memory is a limited capacity system that briefly stores and manages the information required in other cognitive operations [8]. Episodic memory is a past-oriented memory system that encodes, stores, and searches personally experienced events [9].

Existing Interventions for MCI

The increasing prevalence of MCI and the risk of progression to dementia raises questions about interventions that delay or prevent this process [10]. Interventions for MCI can be divided into pharmacological interventions and nonpharmacological interventions. Currently, there are no specific pharmacological interventions for the treatment of MCI. In the United Kingdom, the 2 drugs used for Alzheimer disease, cholinesterase inhibitors and Memantine, have not been shown to help people with MCI [11]. The US Food and Drug Administration has given the drug Aducanumab accelerated approval as a treatment for Alzheimer disease and MCI. However, there is no evidence for the drug's effectiveness data in the treatment of MCI [12]. Most nondrug interventions for MCI address the underlying modifiable causes of MCI, including lifestyle and the treatment of health conditions such as hypertension, obesity, diabetes, stroke, and vitamin deficiency [13]. However, there is no evidence for the effectiveness of dietary changes, including Vitamin E supplements, for delaying MCI [14]. Physical exercise programs have been shown to reduce a person's risk of MCI development [15]. However, the effectiveness of increased physical activity in delaying or delaying the progress of cognitive disorders remains unclear [16]. Other nondrug interventions for MCI include memory training, staying mentally and socially active, and cognitive training [17]. The quality of interventions involving social activities for alleviating MCI remains controversial across existing studies [18,19].

Noncomputerized and Computerized Cognitive Training

With insufficient evidence to support the use of pharmacological and nonpharmacological interventions as described above, cognitive training has been proposed as an intervention to improve cognitive function. This involves repeated activities based on the theory of brain plasticity [20]. With advances in computing technology, traditional cognitive training based on pen and paper has gradually been replaced by computerized cognitive training (CCT) in settings where there is good access to appropriate technology among target groups. CCT is an application of digital health in which individuals can access engaging and interactive cognitive exercises from their own computers, tablets, virtual reality (VR), or mobile devices [21]. CCT involves guided drill-and-practice on standardized tasks, typically without explicit teaching of memory or problem-solving strategies, which distinguish CCT from other approaches for cognitive training [22]. Compared with non-CCT, CCT is more accessible, comprehensive, and flexible to adaptation to individuals' capacity. The game-like nature is often experienced as intrinsically rewarding [23]. In addition, CCT has generated considerable attention as a safe, relatively inexpensive, and scalable intervention that may maintain cognition in older adults [24]. Further, with enjoyable activities, immediate feedback, and automatic adaptations based on participants' performance, CCT is thought to increase participants' motivation and adherence [25].

Existing Studies and Research Gap

Since 2010, a rapidly increasing number of studies started to evaluate the effectiveness of CCT programs specifically targeting certain cognitive domains such as memory [26], executive function, and processing speed [27]. Among them, working memory has garnered particular attention in recent years. A recent systematic review of the effectiveness of CCT has found moderate effect sizes on cognition in healthy older adults [28]. However, the effectiveness of CCT in addressing cognitive decline in people with MCI remains inconclusive. Most of the existing reviews, which synthesized evidences from randomized controlled trials (RCTs) of CCT on participants with MCI, revealed small-to-moderate effects on improving cognitive function [29-32]. Three reviews combined CCT and non-CCT therapies (such as therapeutic drugs, diet modification, and physical activity), providing conclusions about the specific effectiveness of CCT [29,31,32]. A recent Cochrane review included only interventions that lasted more than 12 weeks [33], but that review found only 8 studies with small sample sizes; therefore, conclusions about intervention effectiveness could not be drawn. Considering the rapid development and increasing accessibility of CCT in the last decade, updating the latest evidence about CCT is necessary to inform clinical practice. Therefore, we conducted this review to determine whether CCT is an effective intervention for addressing cognitive decline in

people with MCI. The objectives of this review were to (1) analyze the effectiveness of CCT on preventing progression in cognitive decline and (2) explore the relationship between the characteristics of CCT interventions and cognition-related health outcomes.

Methods

Data Sources and Search Strategy

This systematic review and meta-analysis were performed according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and was registered with PROSPERO (Prospective Register of Systematic Reviews; CRD42021278884). Five web-based databases, that is, MEDLINE, Cochrane, Embase, Web of Science, and Google Scholar were searched and updated in August 2021. The literature search used a combination of search terms and keywords for the following main concepts: “cognitive decline,” “mild cognitive impairment,” “cognitive training,” “cognitive exercise,” “computerized cognitive training,” “virtual reality,” and “technology.” All keywords were concatenated using Boolean operators and appropriate truncation symbols depending on database requirements. The detailed search strategy is shown in Multimedia Appendix 1 [25,34-50]. Snowballing methods identified potential papers by screening reference lists from relevant reviews.

Eligibility Criteria

The inclusion and exclusion criteria were identified based on the PICO (Population, Intervention, Comparison, and Outcomes) approach as follows.

1. Study design: only full-text peer-reviewed RCTs published in English between 2010 and 2021 were included. Pilot studies and studies with abstract only were removed.
2. Population: The population of interest was adults aged 18 years or older who had MCI. Studies including healthy people or those already diagnosed with dementia or with other neurological and psychological disorders were excluded.
3. Intervention: Participants in the experimental groups were treated with CCT only. Studies in which CCT was used along with other therapies or drugs aiming to improve participants’ cognitive functions were removed. The programs used computers, consoles, and VR.
4. Control: Either active control (such as watching general education material and any non–CCT-based training) or usual care (without any intervention applied or waiting list) was included.
5. Outcomes: These included (1) participants’ global cognitive function; (2) specific cognitive function, including executive function, working memory, and episodic memory; and (3) new cases of dementia.

Study Selection and Data Extraction

Two reviewers (RL and RY) independently conducted the initial search of the databases by looking through titles and abstracts. Then, the full text of the included studies was reviewed against the eligibility criteria. The snowballing method was used for the reference lists of the relevant papers. Study citations were imported into the reference management software (Endnote X8.0, Clarivate Analytics) for selection. Any disagreement was resolved by discussing with an additional reviewer.

Three authors extracted the following data: (1) study characteristics (author, year of publication, study location), (2) information of participants (study population, number of patients, gender, age), (3) details about activities in intervention and control group (duration of intervention, frequency of intervention, time per session, delivery device, feedback providing mechanism, interactive patterns, and activities), (4) relevant cognitive function outcomes, including global and specific cognitive function, and (5) when outcomes were measured at multiple time points, measures immediately after the completion of the intervention were extracted. All data were checked by an independent researcher (RL).

Data Synthesis and Analysis

The primary outcome of this review was participants’ global cognitive function, which assessed individuals’ general cognitive status. Secondary outcomes were domain-specific cognitive function, including executive function associated with goal-directed behavior [7]; working memory regarding attentional and short-term memory [8]; episodic memory or long-term memory that encodes, stores, and searches personally experienced events [9]; visual memory; and verbal memory. The R software (R Core Team and the R Foundation for Statistical Computing; version 4.1.2) was used to analyze the quantitative data, and a two-tailed P value of less than .05 was defined as statistically significant. As all the outcomes of effectiveness of CCT were continuous variables, standardized mean differences (SMDs) estimated by Hedge’s g method and their corresponding 95% CIs were used to determine the effect size based on the differences between preintervention and postintervention. For studies with multiple interventions, we calculated the effect size separately for each comparison. Due to the possibility of between-study heterogeneity, the random-effects model was used in the meta-analysis with the pooling method of DerSimonian-Laird. Heterogeneity was evaluated by χ^2 (Cochrane Q), I^2, and Tau^2 statistics and displayed in forest plots. To quantify the magnitude of heterogeneity, we defined a value of I^2 more than 50% as moderate-to-high heterogeneity. Funnel plots were applied to assess publication bias if more than 10 papers were available for an outcome in the meta-analysis. Besides visual inspection, Egger and Begg tests were conducted to adjust the potential effect of publication bias on the interpretation of the results [51]. Furthermore, to test the robustness of the results, sensitivity analyses were conducted using the leave-one-out method. To explore the effects of different characteristics of patients and CCT interventions on the impact of measured effectiveness of global cognitive function, we conducted prespecified subgroup analyses by testing 1 variable at a time. Intervention characteristics included year of publication, delivery devices (computer/tablet or other technology), CCT-targeted domains (multiple or single), feedback provided after treatment or not, interactive patterns (interventions with a patient-provider discussion after treatment), intervention settings (intervention carried out in a group or an individual), and training dose with

cutoff chosen at mean values, including duration (less than 3 months or not), frequency (less than 3 days per week), and time per intervention session (less than 1 hour). Comparator characteristics were defined as whether patterns of activities were actively controlled or passively controlled.

Assessment of Risk of Bias

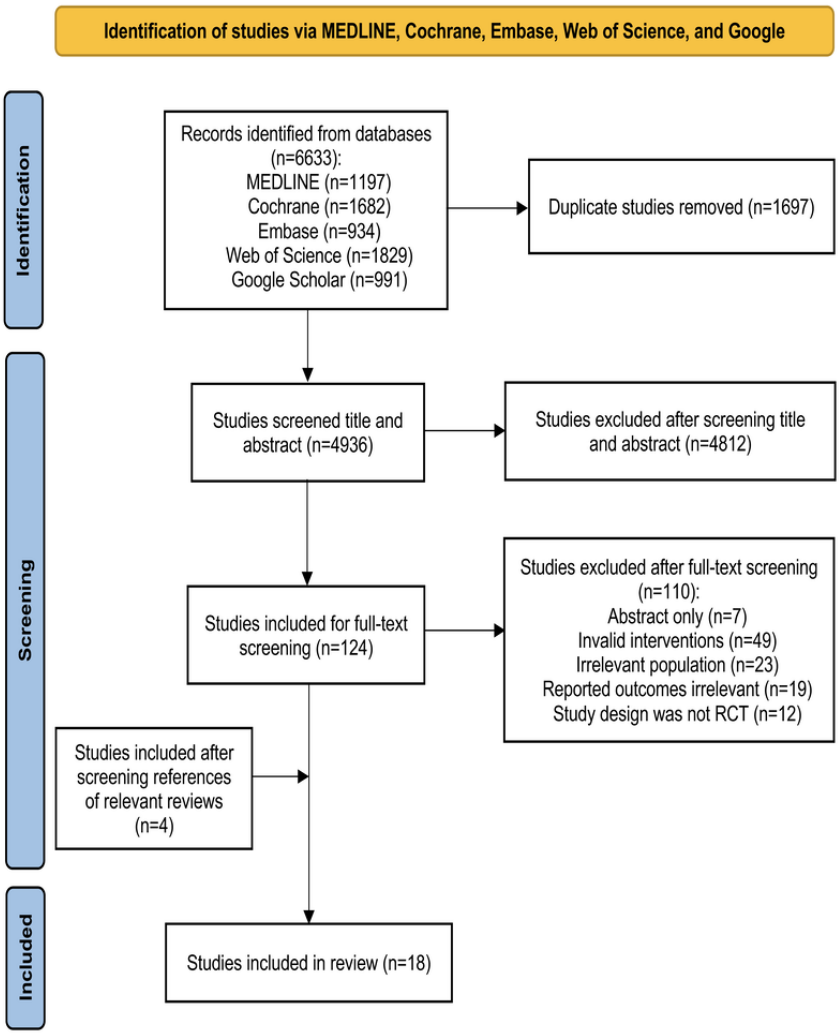
To adequately assess the risk of bias (ROB) in the included studies in this review, the Cochrane ROB tool was used (version 5.4). All information about the features of the process of randomization, allocation concealment, blinding of participants, blinding of outcome assessors, incomplete outcome, and selective reporting were assessed. In addition, the risk of funding bias and baseline imbalance were considered. The ROBs in this review were classified as "high ROB," "low ROB," or "unclear ROB."

Results

Search Results

As shown in Figure 1, the initial search found 4936 studies after excluding 1697 duplicated records. A further 4812 records were excluded after screening the titles and abstracts of the remaining records. A total of 124 full‐text records were assessed for eligibility and 110 records were further excluded. Of these, 7 studies were English abstracts only, 49 studies had invalid interventions (such as the treatment was not CCT or the control group received other interventions with treatment effects), 19 studies reported outcomes irrelevant to the aims of this review (such as safety, acceptance, and feasibility of CCT), 23 studies had irrelevant populations (such as healthy older people and people with dementia), and the study designs of 12 studies were not RCTs. An additional 4 studies were identified from references of relevant reviews. After the above selection process, 18 studies were included in this review.

Figure 1. Flow diagram of study selection. RCT: randomized controlled trial.



Study Characteristics

The characteristics of the participants and trials in the included studies are shown in Table 1. A total of 18 different RCTs with 1059 participants were published between January 2010 and August 2021, 8 of which were published since 2019. Sample sizes ranged from 22 to 141, and the mean age of the participants ranged from 58.8 years to 78.2 years. All studies were conducted in high-and-middle-income countries.

Table 1. Study characteristics.

Study	N[a]	Mean (SD) age (years)	Intervention characteristics	Control group
Liao et al [34], 2020	21/21	73.1 (6.8)	VR[b]: physical activity + cognitive training	Active: combined physical and cognitive training but not CCT[c]-based (reciting poems, crossing obstacles practicing math calculations, etc)
Thapa et al [35], 2020	33/33	72.6 (5.4)	VR: physical activity + cognitive training	Active: educational program on general health care
Park [44], 2020	28/28	71.9 (3.1)	VR: physical activity designed to improve spatial memory	Passive: did not engage in any activity
Park et al [25], 2020	18/17	75.8 (8.5)	VR: spatial cognitive task	Active: tabletop activities, maze and pencil-paper with table activities
Li et al [43], 2019	78/63	69.5 (7.3)	Computer: cognitive training	Passive: not provided with cognitive intervention
Nousia et al [48], 2019	25/21	71.2 (5.1)	Computer: cognitive rehabilitation	Passive: standard clinical care
Yang et al [36], 2019	33/33	75.4 (6.6)	VR: working memory training	Active: reading web-based e-books and playing web-based games such as puzzles
Oh et al [49], 2018	37/16	58.8 (5.0)	Smartphone: cognitive training	Passive: wait-list
Pereira-Morales et al [46], 2017	12/11	64.5 (4.8)	Web application: cognitive training	Active: received only an information brochure to read at home
Savulich et al [37], 2017	21/21	75.2 (7.4)	iPad game: cognitive training	Passive: clinic as usual
Han et al [47], 2017	43/42	73.7 (4.8)	iPad tablet: cognitive training	Passive: usual care
Hyer et al [50], 2016	34/34	75.1 (7.4)	Computer: working memory training	Active: sham cognitive training
Gooding et al [41], 2016 (CCT), and Gooding et al [41], 2016 (CVT[d])	31/20, 23/20	75.6 (8.8)	Computer: plasticity-based training program; Computer: traditional CCT that is embedded within Neuropsychological and Educational Approach to Remediation model of treatment	Active: computer games and puzzles
Barban et al [39], 2016	46/60	74.4 (5.7)	Computer: reminiscence therapy + cognitive training	Passive: cross-over (rest)
Styliadis et al [42], 2015	14/14	67.6 (4.0)	Computer: cognitive training	Active: underwent a training protocol consisting of watching a documentary and answering questionnaire
Fiatarone Singh et al [40], 2014	22/24	70.1 (6.7)	Computer: cognitive training + sham exercise	Active: sham cognitive (watch short videos) + sham exercise (stretching and seated calisthenics)
Bozoki et al [45], 2013	32/28	68.9 (6.8)	Computer: cognitive training + mental training	Active: thoughts in motion; sound thinking; headline clues
Herrera et al [38], 2012	11/11	78.2 (1.4)	Computer: memory and attention training	Active: cognitive activities including find names of countries and corresponding capitals etc

[a]Number of participants in intervention group/control group.

[b]VR: virtual reality.

[c]CCT: computerized cognitive training.

[d]CVT: cognitive vitality training.

CCT Characteristics

Common activities included attention training, visual processing, sensory integration, and recollection exercises. Thirteen studies were delivered as cognitive training programs on computers or tablets [37-43,45-50]. Another 5 studies [25,34-36,44] used VR-based interactive video games, with 1 study combining both tablets and VR devices [37]. In 5 studies, participants completed all treatment in groups under supervision by trained cognitive therapists [27,37-40]. Others carried out CCT interventions by themselves. The frequency of CCT sessions was 2-5 times per week, with a mean frequency of 3 times per week. The length of each session was around an hour in all 18 studies. Mean trial duration was 10.5 (range 4-24) weeks. The average dropout rate in the studies was 8% (range 0%-23%). The main reasons for dropout were unwillingness to continue and unrelated health

issues. Eight studies reported no missing data from baseline to completion [25,36-38,42,44,46,48]. The activities of the CCT programs were diverse and 7 of them targeted multidomain cognitive function [34,35,39-43]. Most CCT programs included more than 1 activity, including remembering items in a limited time, mathematical calculations, and auditory stimuli (an auditory stimulus and recognizing a synthetically generated syllable from a confusable pair). CCT interventions in some studies, especially VR-based CCT interventions, inevitably combined some physical activities [34,35,39] such as balance training, agility training, strength training, and flexibility training. In 7 studies [25,34,37,41,44-46], feedback was provided to the participants, either in real time or as they finished each activity during the CCT session, such as "Good job," "Better next time," and visual and auditory feedback. Seven studies conducted interactions between providers and patients in CCT groups during the intervention or after they finished each session [25,35,38-40,46,47].

Outcome Measures

Global Cognitive Function

Eleven studies measured the change in global cognitive function between preintervention and postintervention immediately after completion of the whole treatment by using Mini-Mental State Examination [35-37,39,41-43,47], Montreal Cognitive Assessment [25,34], or Alzheimer's Disease Assessment Scale-Cognitive subscale [40]. Gooding et al [41] had more than one intervention group with the same outcomes measured. Therefore, 12 trials were shown in the meta-analysis of global cognitive function. The pooled SMD of global cognitive function (Figure 2) showed a statistically significant improvement for participants in the intervention groups compared to that in the control groups (SMD=0.54, 95% CI 0.35-0.73), with moderate heterogeneity between studies (P=.09; I^2=38%). No significant publication bias was suggested, as no asymmetry was detected in the funnel plot (Multimedia Appendix 2), and neither Egger (P=.98) nor Begg (P=.89) tests were significant. Effect size in sensitivity analysis remained significant with no notable change (Multimedia Appendix 3 [25,34-37,39-43,47]). Subgroup analyses (Table 2) showed consistent improvement in global cognitive behavior in the CCT intervention groups across all variables mentioned above. However, we observed no significant difference in the effect size in each comparison.

Figure 2. Forest plot for global cognitive function [25,34-37,39-43,47]. CCT: computerized cognitive training; CVT: cognitive vitality training; SMD: standardized mean difference.

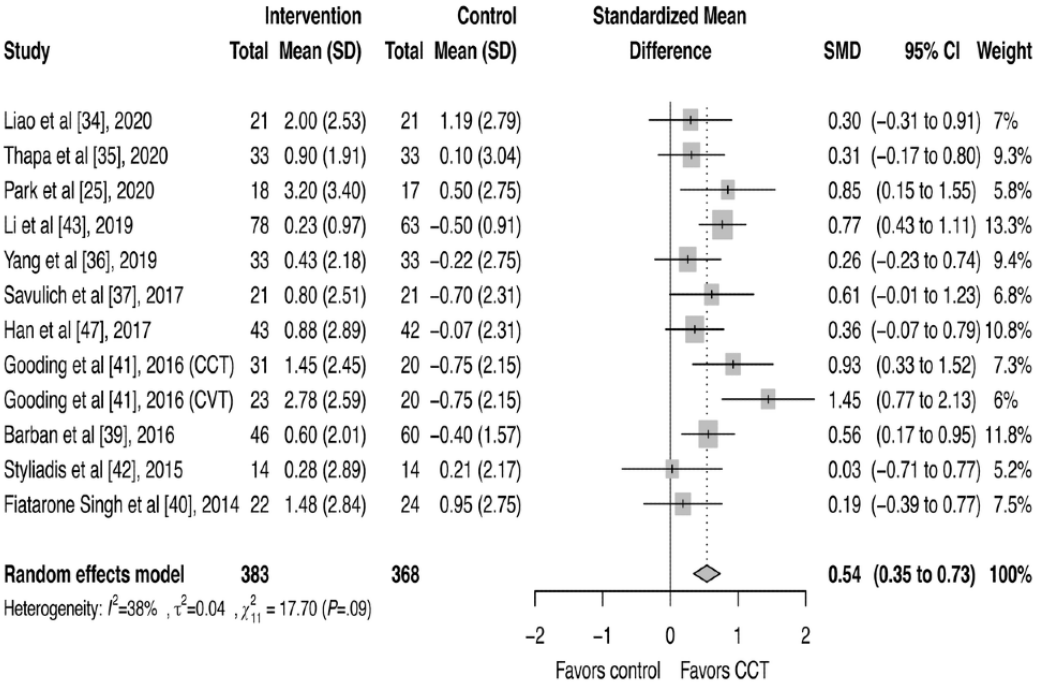


Table 2. Subgroup analyses of global cognitive function.

Study subgroup characteristic	Trials (n)	I^2 (%)	Standardized mean difference (95% CI)	*P* value
Overall	12	38	0.54 (0.35-0.73)	N/A[a]
Intervention characteristics				
Publication[b]				.75
2019 or later	5	23	0.51 (0.26-0.76)	
Prior to 2019	7	52	0.58 (0.27-0.88)	
Delivery devices				.22
Computer/tablets	8	48	0.61 (0.35-0.86)	
Virtual reality	4	0	0.38 (0.10-0.65)	
Computerized cognitive training type				.51
Multidomain	8	52	0.57 (0.30-0.84)	
Single domain	4	0	0.45 (0.18- 0.71)	
Interaction				.33
With	5	0	0.43 (0.22-0.65)	
Without	7	54	0.62 (0.31-0.93)	
Feedback				.29
With	3	67	0.85 (0.19-1.52)	
Without	9	13	0.49 (0.31-0.66)	
Setting				.96
Activities in group with supervision	4	0	0.53 (0.27-0.80)	
Individual	8	55	0.54 (0.27-0.82)	
Training dose				
Duration				.10
>3 months	4	62	0.81 (0.37-1.24)	
≤3 months	8	0	0.41 (0.23-0.60)	
Frequency				.39
≥3 days per week	6	27	0.46 (0.21-0.71)	
<3 days per week	6	52	0.64 (0.33-0.95)	
Time				.96
≥1 h per session	8	47	0.54 (0.26-0.82)	
<1 h per session	4	32	0.55 (0.27-0.83)	
Comparator characteristics				.69
Active control	8	54	0.52 (0.21-0.83)	
Inactive control	4	0	0.59 (0.39-0.80)	

[a]N/A: not applicable.

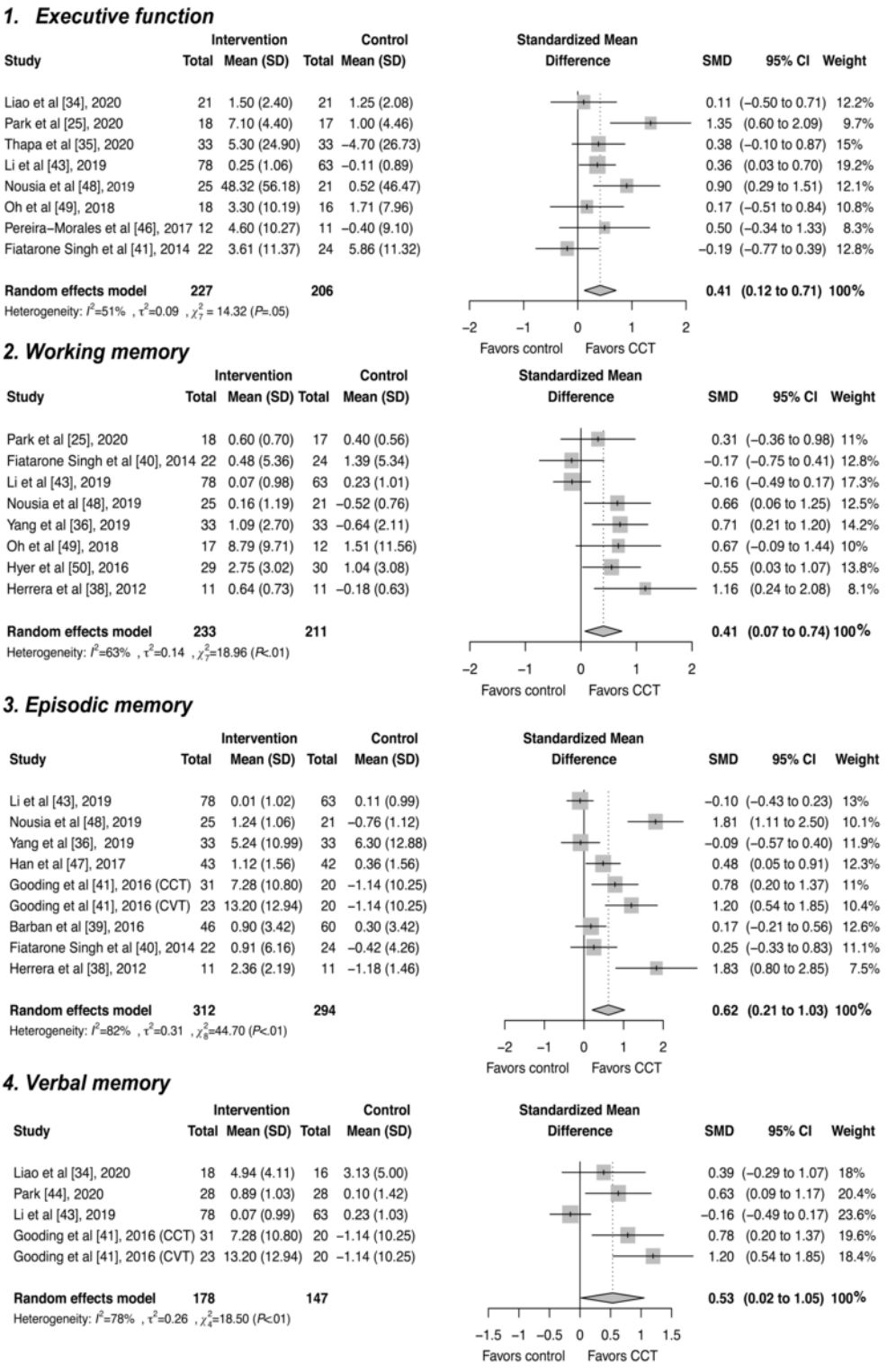
[b]Cutoff chosen was the year this updated review added newly published studies compared with the latest published review [30].

Domain-Specific Cognition

Executive Function

Eight studies assessed the change of executive function by using the Stroop test [43,46], the Trail Making Test [25,35,48], Controlled Oral Word Association Test [40], Memory Diagnostic System (executive subscale) [49], and Executive Interview [34]. As shown in Figure 3, the overall pooled SMD of executive function was 0.41 (95% CI 0.12-0.71), with moderate inconsistency between the studies (P=.046; I^2=51%), but no publication bias was presented (Multimedia Appendix 2). The sensitivity analysis provided results consistent with the original result (Multimedia Appendix 3).

Figure 3. Forest plot for studies assessing specific domains of cognitive function (executive function, working memory, episodic memory, and verbal memory) [25,34-36,38-41,43,44,46-50]. CCT: computerized cognitive training; CVT: cognitive vitality training; SMD: standardized mean difference.



Working Memory

A total of 8 studies measured the change in the working memory. The digit span test was the most common instrument used to measure this outcome [25,36,38,48], followed by Auditory Logical Memory and Auditory Verbal Logical Test (immediate recall) [40,43], memory diagnostic system (working memory subscale) [49], and Span Board [50]. The working memory of the participants in the intervention groups showed an improvement compared to that of those in the control groups (SMD=0.41, 95% CI 0.07-0.74) (Figure 3). Heterogeneity across the studies was moderate (P=.008; I^2=63%) (Multimedia Appendix 2).

Episodic Memory

A total of 8 studies measured the change in episodic memory by varied delayed memory recall tests [36,38-41,43,47,48]. The

forest plot for episodic memory is presented in Figure 3, with a pooled SMD of 0.62 (95% CI 0.21-1.03), reflecting the benefit from the intervention group. However, the heterogeneity analyses suggested considerable heterogeneity between the studies (P<.001; I^2=82%) (Multimedia Appendix 2).

Verbal and Visual Memory

Four studies specifically investigated the change in participants' verbal memory [34,41,43,44] and revealed an SMD of 0.53 (95% CI 0.02-1.05; P=.001; I^2=78%) in favor of CCT groups (Figure 3). No publication bias was detected in the funnel plot (Multimedia Appendix 2). In terms of the visual memory, only 1 study measured visual memory based on the Wechsler Memory Scale-Revised Visual Reproduction subset [41], but there was no significant difference between the CCT group and control (SMD=0.33, 95% CI –0.08 to 0.75).

Other Findings

Adverse Effects

There were no adverse events reported from the CCT interventions across the 18 studies. However, the study conducted by Fiatarone Singh et al [40] revealed 2 adverse events in the control groups due to falls or pre-existing arthritis symptoms exacerbated while participating in strength testing or training.

Effect Durability and Feasibility

Five studies reported additional assessments after the end of interventions [38-40,43,50]. The duration of the follow‐up after the end of the interventions ranged from 3 to 12 months. All 5 studies evaluated the long-term maintenance of CCT-related cognitive benefits. Out of the various cognitive measures, all reported some sustained improvement, significantly better than controls. Notably, only 1 study reported dementia incidence after the training [43]. Three of the total 78 patients in the CCT group were diagnosed with Alzheimer disease in 6 months and another 3 (33 assessed) developed Alzheimer disease over 12 months after cessation of training, compared with 15 out of 63 and 6 out of 30 in the control group, respectively. No study measured participants' satisfaction pertaining to the intervention itself. However, improved overall memory satisfaction and psychosocial satisfaction were reported [40,49].

ROB With Studies

As depicted in Figure 4 and Figure 5, no study exhibited a low ROB in all items of assessment, while 9 studies had a high ROB in at least one item of assessment [25,34,38,39,42,44,45,49,50]. Overall, 11 studies were assessed as low risk of selection bias [25,34-37,39,40,44,46,47,49], and another 7 studies [38,41-43,45,48,50] were assessed as unclear because they did not report a clear process of generation of a randomized sequence. Four studies had a high risk of performance bias, as participants were unmasked during the treatment [34,38,39,44]. The risk of detection bias was high in 3 studies [42,45,50], as outcome assessors were not blinded to the intervention allocation. Other biases were judged as low risk in all 18 studies.

Figure 4. Results of risk of bias presented as percentages across all included studies.

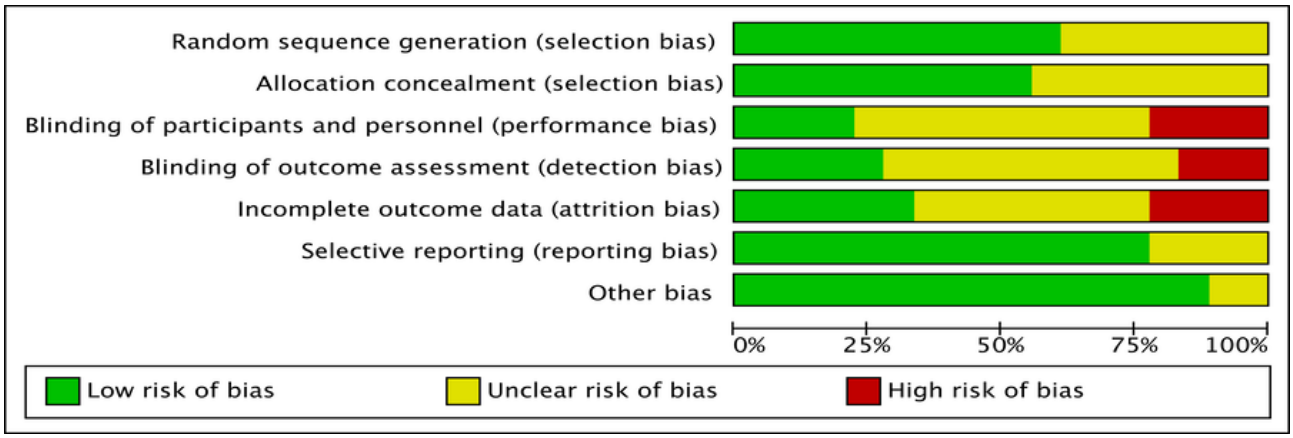
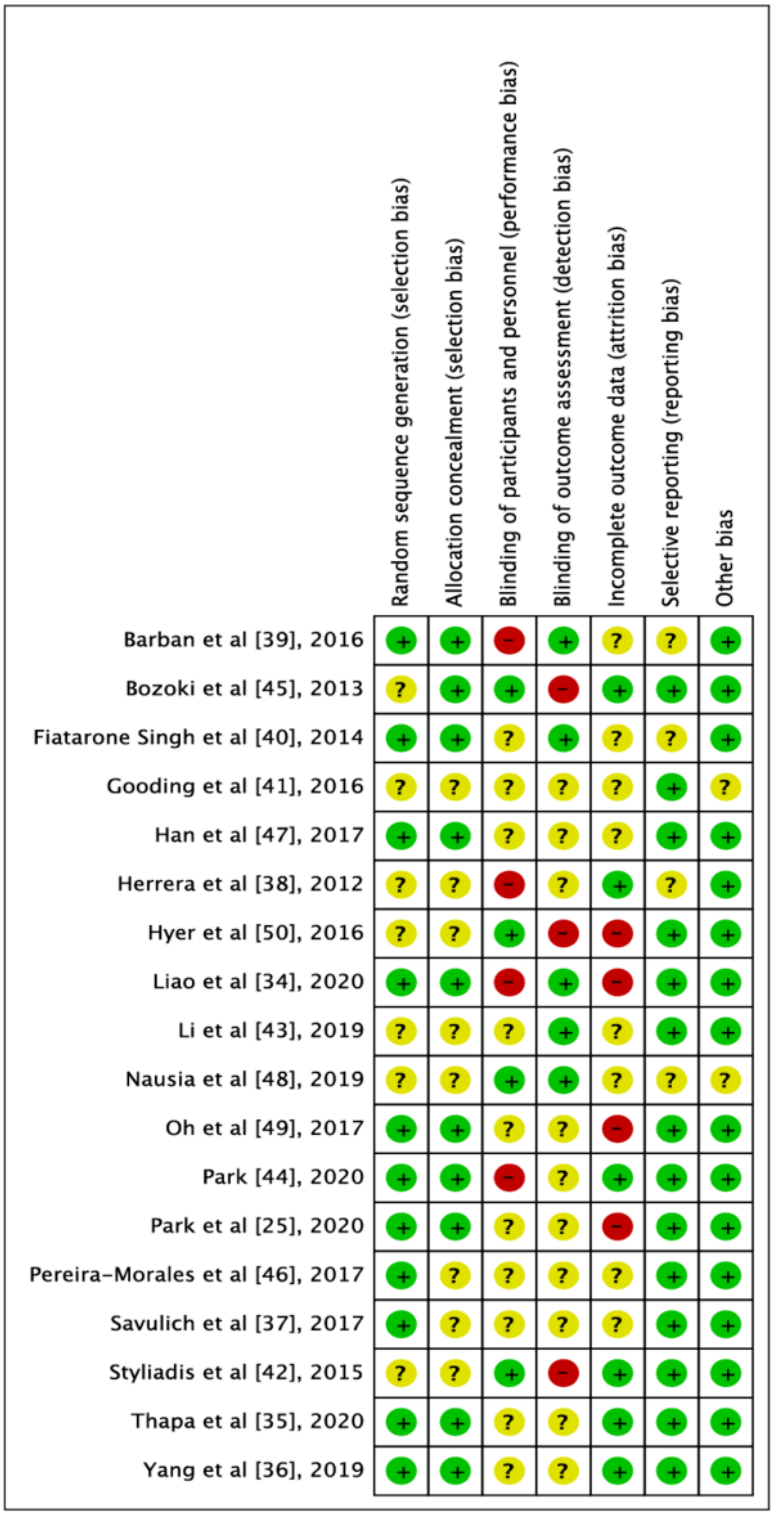


Figure 5. Results of risk of bias for each included study [25,34-50]. Key for colors: Red: high ROB; Yellow: unclear ROB; Green: low ROB. ROB: risk of bias.



Discussion

Summary of the Principal Findings

This systematic review synthesized 18 RCTs with a total of 1059 participants to assess the effectiveness of CCT in delaying the progression of MCI. The findings of this review indicated that CCT interventions provided a statistically significant improvement in global cognitive function. In addition, CCT interventions resulted in a positive effect in executive function, working memory, episodic memory, and verbal memory in people with cognitive decline compared to those in the control groups. We analyzed the relationship between the characteristics of CCT interventions and cognition-related health outcomes by using meta-analyses. Our results emphasized that CCT is a promising approach for improving global cognitive function. According to the subgroup analyses, more effective interventions were those that were performed within patients' groups, which used interaction and feedback between providers and patients,

and those targeting multidomain cognitive functions with longer durations per course and longer sessions, although the effect size is marginal and is not statistically significant. Interestingly, although not reaching a level of significance, subgroup analysis showed that the effect sizes in studies involving CCT sessions with no more than 3 times per week appeared to be higher than those in studies involving CCT sessions more than 3 times per week. This finding is in line with a previous meta-analysis, which showed that the intensive frequency of CCT sessions resulted in worse outcomes and training fatigue [51]. Therefore, future research should include variations in frequency of CCT delivery to assess the impact of the different treatment doses of CCT and to determine the frequency of the sessions with optimal outcomes. Although we categorized studies by all essential characteristics of CCT interventions mentioned from previous reviews, we did not identify any specific characteristics that could improve the effect of the CCT in global function. Possible explanations include the strict inclusion criteria, which meant there were a small number of studies in each subgroup as well as marked heterogeneity in study design, and the effect might be confounded by other factors that were not identified.

We found that those interventions conducted after 2019 [25,34-36] were more likely to deliver CCT by VR. VR-based training could help to overcome the barriers of lack of infrastructure, enhance motivation, and increase user participation by resembling real-life scenarios [52]. In addition, physical activities were often added to CCT interventions, especially in VR-delivered cognitive sessions. In future studies, it would be important to investigate the beneficial or synergistic effects of the combination of cognitive and physical components, especially for using such applications among older adults with MCI [53]. No study on the cost-effectiveness of CCT of MCI has been conducted. However, economic analysis is necessary for further research, especially given the huge economic burden of dementia for the society and family. Compared with traditional cognitive training, CCT is largely web-based, facilitating dissemination, and not requiring highly trained cognitive trainers (as does traditional cognitive training), and this considerably reduces the cost for patients and health systems.

Limitations of This Study

The chief weakness of this review is the small number of studies included, especially at the level of subgroup. Second, this systematic review only included English language studies published in peer-reviewed journals, thereby potentially reducing the diversity of studies. Third, the mean age range of the participants in our studies was 68-76 years. Petersen et al [5] found that the risk of older people aged 80-84 years developing MCI is almost 4 times higher than that of those aged 60-64 years; therefore, we may lack data on the age group that might most benefit from the interventions.

Implications for Further Study

Our review clearly shows that the quality of evidence is overall low, with small sample sizes, short follow-up duration, and imbalanced number of studies with different CCT characteristics, especially at the subgroup level. Although we show overall statistical significance, clinical significance is still questionable, and there is insufficient evidence to support the scale-up of such treatments. Several suggestions to improve the quality of trials are as follows.

First, longer term follow-up is needed. Only 4 studies conducted follow-up assessments after the end of the interventions [38-40,51]. The number of participants who develop dementia during the follow-up should also be an outcome measure in further studies. Further, concerning the problem of study design, the sample size of the included studies was small, ranging from 22 to 141. In addition, some intervention and control activities were similar, and this might have counteracted the effect of the CCT. Although computerized programs allow cognitive training designed to target specific cognitive capabilities, the problem of transfer of effects to tasks and cognitive domains not directly trained is a major issue in CCT [54]. Therefore, future research should clearly differentiate CCT interventions and control groups and identify the effectiveness of specific cognitive capacities. Further, more studies call for comprehensive analyses of the effectiveness of dual-task approaches such as cognitive training accompanied with physical activities.

Second, concerning statistical analysis, no power calculation was conducted in the included studies. It is important for studies to present sample size calculations to improve the validity of the results [55]. In addition, if the achieved smaller size differs from the planned sample size, the limitations for the implications need to be addressed.

Third, to date, there are no well-established CCT treatment guidelines. Most of the activities in the interventions were designed without standard criteria, including technical details, feasibility, and sustainability of the intervention strategies. The evidence in this review is heterogeneous in quality, completeness, and objectivity of the reporting of CCT interventions, thus making comparisons across intervention activities difficult. This is partly attributable to the multidisciplinary nature of CCT, which combines different approaches from the fields of health care and technology. The rapid pace of CCT development often outpaces the research ability to generate evidence. Therefore, a set of standards is needed, which can harmonize and improve the quality of CCT intervention, both for implementation and evidence dissemination.

Conclusions

With aging populations increasing globally, there is a huge interest in interventions to delay or prevent cognitive decline. The findings from this review suggest that CCT may be a promising approach to improve global cognitive function and executive function. High accessibility and no necessity for delivery by trained experts are the major advantages of CCT as a clinical tool. However, studies with rational sample sizes, long-term treatment, and sufficient follow-up duration are needed to provide the evidence for recommendations for integration into clinical practice.

Acknowledgments

This study was funded by the Centre of Global Health, Zhejiang University. The authors would like to thank Ranghao Bao for proofreading and RY's family for their spiritual support, particularly his grandmother, a brave and optimistic old lady with Alzheimer disease.

Data Availability

All data analyzed during this study are from published papers in web-based databases and included in this paper. The data sets generated during and analyzed in our study are available from the corresponding author on reasonable request.

Authors' Contributions

RL and JG contributed to data search, extraction, and analysis. RL and RY drafted and revised this paper. YG contributed to data extraction. JG advised on data analysis and revised the paper. TH initiated the research, revised the paper, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and additional study characteristics [25,34-50].
[DOCX File , 24 KB - jmir_v24i10e38624_app1.docx]

Multimedia Appendix 2
Funnel plot.
[DOCX File , 105 KB - jmir_v24i10e38624_app2.docx]

Multimedia Appendix 3
Sensitivity analyses [25,34-37,39-43,46-49].
[DOCX File , 378 KB - jmir_v24i10e38624_app3.docx]

References

1. Riello M, Rusconi E, Treccani B. The Role of Brief Global Cognitive Tests and Neuropsychological Expertise in the Detection and Differential Diagnosis of Dementia. Front Aging Neurosci 2021;13:648310 [FREE Full text] [doi: 10.3389/fnagi.2021.648310] [Medline: 34177551]
2. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Jan;18(1):88-106 [FREE Full text] [doi: 10.1016/S1474-4422(18)30403-4] [Medline: 30497964]
3. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 2018 Jan 16;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
4. Thaipisuttikul P, Jaikla K, Satthong S, Wisajun P. Rate of conversion from mild cognitive impairment to dementia in a Thai hospital-based population: A retrospective cohort. Alzheimers Dement (N Y) 2022;8(1):e12272 [FREE Full text] [doi: 10.1002/trc2.12272] [Medline: 35386122]
5. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med 2014 Mar;275(3):214-228 [FREE Full text] [doi: 10.1111/joim.12190] [Medline: 24605806]
6. Palmqvist S, Janelidze S, Quiroz YT, Zetterberg H, Lopera F, Stomrud E, et al. Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. JAMA 2020 Aug 25;324(8):772-781 [FREE Full text] [doi: 10.1001/jama.2020.12134] [Medline: 32722745]
7. Buss A, Lowery K. Inhibitory control and executive function. In: Benson JB, editor. Encyclopedia of Infant and Early Childhood Development. Oxford: Elsevier; Mar 03, 2020.
8. Camos V, Johnson M, Loaiza V, Portrat S, Souza A, Vergauwe E. What is attentional refreshing in working memory? Ann N Y Acad Sci 2018 Jul;1424(1):19-32. [doi: 10.1111/nyas.13616] [Medline: 29542133]
9. Stark S, Stark C. Chapter 67 - introduction to memory. In: Hickok G, Small SL, editors. Neurobiology of Language. San Diego: Academic Press; Aug 15, 2015.
10. Dementia and risk reduction: An analysis of protective and modifiable risk factors. Alzheimer's Disease International. 2014. URL: https://www.alzint.org/u/WorldAlzheimerReport2014.pdf [accessed 2021-10-17]

11. Birks J. Cholinesterase inhibitors for Alzheimer's disease. Cochrane Database Syst Rev 2006 Jan 25(1):CD005593 [FREE Full text] [doi: 10.1002/14651858.CD005593] [Medline: 16437532]
12. Herring WL, Gould IG, Fillit H, Lindgren P, Forrestal F, Thompson R, et al. Predicted Lifetime Health Outcomes for Aducanumab in Patients with Early Alzheimer's Disease. Neurol Ther 2021 Dec;10(2):919-940 [FREE Full text] [doi: 10.1007/s40120-021-00273-0] [Medline: 34426940]
13. Oltra-Cucarella J, Ferrer-Cascales R, Alegret M, Gasparini R, Díaz-Ortiz LM, Ríos R, et al. Risk of progression to Alzheimer's disease for different neuropsychological Mild Cognitive Impairment subtypes: A hierarchical meta-analysis of longitudinal studies. Psychol Aging 2018 Nov;33(7):1007-1021. [doi: 10.1037/pag0000294] [Medline: 30284855]
14. Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and Donepezil for the Treatment of Mild Cognitive Impairment. N Engl J Med 2005 Jun 09;352(23):2379-2388. [doi: 10.1056/nejmoa050151]
15. Ribeiro IC, Teixeira CVDL, Rezende TJR, Magalhaes TNC, Cendes F, Balthazar M. Effects of 6 months of multicomponent exercise on DTI parameters of MCI in the AD continuum. Alzheimer's & Dementia 2022 Feb 01;17(S7):e052539. [doi: 10.1002/alz.052539]
16. Huang P, Fang R, Li B, Chen S. Exercise-Related Changes of Networks in Aging and Mild Cognitive Impairment Brain. Front Aging Neurosci 2016;8:47 [FREE Full text] [doi: 10.3389/fnagi.2016.00047] [Medline: 27014055]
17. Stott J, Spector A. A review of the effectiveness of memory interventions in mild cognitive impairment (MCI). Int Psychogeriatr 2011 May;23(4):526-538. [doi: 10.1017/S1041610210001973] [Medline: 20946704]
18. Hughes TF, Flatt JD, Fu B, Chang CH, Ganguli M. Engagement in social activities and progression from mild to severe cognitive impairment: the MYHAT study. Int Psychogeriatr 2013 Apr;25(4):587-595 [FREE Full text] [doi: 10.1017/S1041610212002086] [Medline: 23257280]
19. James BD, Wilson RS, Barnes LL, Bennett DA. Late-life social activity and cognitive decline in old age. J Int Neuropsychol Soc 2011 Nov;17(6):998-1005 [FREE Full text] [doi: 10.1017/S1355617711000531] [Medline: 22040898]
20. Li H, Li J, Li N, Li B, Wang P, Zhou T. Cognitive intervention for persons with mild cognitive impairment: A meta-analysis. Ageing Res Rev 2011 Apr;10(2):285-296. [doi: 10.1016/j.arr.2010.11.003] [Medline: 21130185]
21. A consensus on the brain training industry from the scientific community (full statement). Stanford Center on Longevity. 2014. URL: https://longevity.stanford.edu/a-consensus-on-the-brain-training-industry-from-the-scientific-community-2/ [accessed 2021-10-17]
22. Bahar-Fuchs A, Martyr A, Goh AM, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. Cochrane Database Syst Rev 2019 Mar 25;3:CD013069 [FREE Full text] [doi: 10.1002/14651858.CD013069.pub2] [Medline: 30909318]
23. Caggianese G, Chirico A, De PG, Gallo L, Giordano A, Predazzi M. Towards a virtual reality cognitive training system for mild cognitive impairment and Alzheimer's disease patients. 2018 Presented at: 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA); May 18; Caserta, Italy p. 663-667. [doi: 10.1109/waina.2018.00164]
24. Liang J, Xu Y, Lin L, Jia R, Zhang H, Hang L. Comparison of multiple interventions for older adults with Alzheimer disease or mild cognitive impairment: A PRISMA-compliant network meta-analysis. Medicine (Baltimore) 2018 May;97(20):e10744 [FREE Full text] [doi: 10.1097/MD.0000000000010744] [Medline: 29768349]
25. Park J, Jung Y, Lee G. Virtual Reality-Based Cognitive-Motor Rehabilitation in Older Adults with Mild Cognitive Impairment: A Randomized Controlled Study on Motivation and Cognitive Function. Healthcare (Basel) 2020 Sep 11;8(3):335 [FREE Full text] [doi: 10.3390/healthcare8030335] [Medline: 32932997]
26. Niu H, Li X, Chen Y, Ma C, Zhang J, Zhang Z. Reduced frontal activation during a working memory task in mild cognitive impairment: a non-invasive near-infrared spectroscopy study. CNS Neurosci Ther 2013 Feb;19(2):125-131 [FREE Full text] [doi: 10.1111/cns.12046] [Medline: 23279823]
27. Guarino A, Forte G, Giovannoli J, Casagrande M. Executive functions in the elderly with mild cognitive impairment: a systematic review on motor and cognitive inhibition, conflict control and cognitive flexibility. Aging Ment Health 2020 Jul;24(7):1028-1045. [doi: 10.1080/13607863.2019.1584785] [Medline: 30938193]
28. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
29. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. Am J Psychiatry 2017 Apr 01;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
30. Zhang H, Huntley J, Bhome R, Holmes B, Cahill J, Gould RL, et al. Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: a systematic review and meta-analysis. BMJ Open 2019 Aug 18;9(8):e027062 [FREE Full text] [doi: 10.1136/bmjopen-2018-027062] [Medline: 31427316]
31. Ge S, Zhu Z, Wu B, McConnell ES. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018 Sep 15;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]

32. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
33. Gates NJ, Vernooij RW, Di Nisio M, Karim S, March E, Martínez G, et al. Computerised cognitive training for preventing dementia in people with mild cognitive impairment. Cochrane Database Syst Rev 2019 Mar 13;3:CD012279 [FREE Full text] [doi: 10.1002/14651858.CD012279.pub2] [Medline: 30864747]
34. Liao Y, Tseng H, Lin Y, Wang C, Hsu W. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020 Feb;56(1):47-57 [FREE Full text] [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
35. Thapa N, Park HJ, Yang J, Son H, Jang M, Lee J, et al. The Effect of a Virtual Reality-Based Intervention Program on Cognition in Older Adults with Mild Cognitive Impairment: A Randomized Control Trial. J Clin Med 2020 Apr 29;9(5):1283 [FREE Full text] [doi: 10.3390/jcm9051283] [Medline: 32365533]
36. Yang H, Chu H, Kao C, Chiu H, Tseng I, Tseng P, et al. Development and effectiveness of virtual interactive working memory training for older people with mild cognitive impairment: a single-blind randomised controlled trial. Age Ageing 2019 Jul 01;48(4):519-525. [doi: 10.1093/ageing/afz029] [Medline: 30989165]
37. Savulich G, Piercy T, Fox C, Suckling J, Rowe JB, O'Brien JT, et al. Cognitive Training Using a Novel Memory Game on an iPad in Patients with Amnestic Mild Cognitive Impairment (aMCI). Int J Neuropsychopharmacol 2017 Aug 01;20(8):624-633 [FREE Full text] [doi: 10.1093/ijnp/pyx040] [Medline: 28898959]
38. Herrera C, Chambon C, Michel BF, Paban V, Alescio-Lautier B. Positive effects of computer-based cognitive training in adults with mild cognitive impairment. Neuropsychologia 2012 Jul;50(8):1871-1881. [doi: 10.1016/j.neuropsychologia.2012.04.012] [Medline: 22525705]
39. Barban F, Annicchiarico R, Pantelopoulos S, Federici A, Perri R, Fadda L, et al. Protecting cognition from aging and Alzheimer's disease: a computerized cognitive training combined with reminiscence therapy. Int J Geriatr Psychiatry 2016 Apr;31(4):340-348. [doi: 10.1002/gps.4328] [Medline: 26205305]
40. Fiatarone Singh MA, Gates N, Saigal N, Wilson GC, Meiklejohn J, Brodaty H, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. J Am Med Dir Assoc 2014 Dec;15(12):873-880. [doi: 10.1016/j.jamda.2014.09.010] [Medline: 25444575]
41. Gooding AL, Choi J, Fiszdon JM, Wilkins K, Kirwin PD, van Dyck CH, et al. Comparing three methods of computerised cognitive training for older adults with subclinical cognitive decline. Neuropsychol Rehabil 2016 Oct;26(5-6):810-821. [doi: 10.1080/09602011.2015.1118389] [Medline: 26674122]
42. Styliadis C, Kartsidis P, Paraskevopoulos E, Ioannides AA, Bamidis PD. Neuroplastic effects of combined computerized physical and cognitive training in elderly individuals at risk for dementia: an eLORETA controlled study on resting states. Neural Plast 2015;2015:172192 [FREE Full text] [doi: 10.1155/2015/172192] [Medline: 25945260]
43. Li B, He N, Qiao Y, Xu H, Lu Y, Cui P, et al. Computerized cognitive training for Chinese mild cognitive impairment patients: A neuropsychological and fMRI study. Neuroimage Clin 2019;22:101691 [FREE Full text] [doi: 10.1016/j.nicl.2019.101691] [Medline: 30708349]
44. Park J. Effects of virtual reality-based spatial cognitive training on hippocampal function of older adults with mild cognitive impairment. Int Psychogeriatr 2022 Feb;34(2):157-163. [doi: 10.1017/S1041610220001131] [Medline: 32616109]
45. Bozoki A, Radovanovic M, Winn B, Heeter C, Anthony JC. Effects of a computer-based cognitive exercise program on age-related cognitive decline. Arch Gerontol Geriatr 2013;57(1):1-7. [doi: 10.1016/j.archger.2013.02.009] [Medline: 23542053]
46. Pereira-Morales AJ, Cruz-Salinas AF, Aponte J, Pereira-Manrique F. Efficacy of a computer-based cognitive training program in older people with subjective memory complaints: a randomized study. Int J Neurosci 2018 Jan;128(1):1-9. [doi: 10.1080/00207454.2017.1308930] [Medline: 28316267]
47. Han JW, Son KL, Byun HJ, Ko JW, Kim K, Hong JW, et al. Efficacy of the Ubiquitous Spaced Retrieval-based Memory Advancement and Rehabilitation Training (USMART) program among patients with mild cognitive impairment: a randomized controlled crossover trial. Alzheimers Res Ther 2017 Jun 06;9(1):39 [FREE Full text] [doi: 10.1186/s13195-017-0264-8] [Medline: 28587629]
48. Nousia A, Martzoukou M, Siokas V, Aretouli E, Aloizou A, Folia V, et al. Beneficial effect of computer-based multidomain cognitive training in patients with mild cognitive impairment. Appl Neuropsychol Adult 2021;28(6):717-726. [doi: 10.1080/23279095.2019.1692842] [Medline: 31885287]
49. Oh SJ, Seo S, Lee JH, Song MJ, Shin M. Effects of smartphone-based memory training for older adults with subjective memory complaints: a randomized controlled trial. Aging Ment Health 2018 Apr;22(4):526-534. [doi: 10.1080/13607863.2016.1274373] [Medline: 28071929]
50. Hyer L, Scott C, Atkinson MM, Mullen CM, Lee A, Johnson A, et al. Cognitive Training Program to Improve Working Memory in Older Adults with MCI. Clin Gerontol 2016;39(5):410-427. [doi: 10.1080/07317115.2015.1120257] [Medline: 29471774]

51. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
52. Strojny P, Traczyk Z, Rozycka M, Bem W, Sawicki W. Fourier analysis of nuclear and cytoplasmic shape of blood lymphoid cells from healthy donors and chronic lymphocytic leukemia patients. Anal Quant Cytol Histol 1987 Dec;9(6):475-479. [Medline: 3435625]
53. Berridge C, Turner NR, Liu L, Karras SW, Chen A, Fredriksen-Goldsen K, et al. Advance Planning for Technology Use in Dementia Care: Development, Design, and Feasibility of a Novel Self-administered Decision-Making Tool. JMIR Aging 2022 Jul 27;5(3):e39335 [FREE Full text] [doi: 10.2196/39335] [Medline: 35896014]
54. Harvey PD, McGurk SR, Mahncke H, Wykes T. Controversies in Computerized Cognitive Training. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Nov;3(11):907-915 [FREE Full text] [doi: 10.1016/j.bpsc.2018.06.008] [Medline: 30197048]
55. Li R, Liang N, Bu F, Hesketh T. The Effectiveness of Self-Management of Hypertension in Adults Using Mobile Health: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]

Abbreviations

CCT: computerized cognitive training
MCI: mild cognitive impairment
PICO: Population, Intervention, Comparison, and Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
RCT: randomized controlled trial
ROB: risk of bias
SMD: standardized mean difference
VR: virtual reality

Edited by T Leung; submitted 10.04.22; peer-reviewed by N Chalghaf, G Binarelli; comments to author 19.05.22; revised version received 05.09.22; accepted 22.09.22; published 27.10.22.

Please cite as:
Li R, Geng J, Yang R, Ge Y, Hesketh T
Effectiveness of Computerized Cognitive Training in Delaying Cognitive Function Decline in People With Mild Cognitive Impairment: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(10):e38624
URL: https://www.jmir.org/2022/10/e38624
doi:10.2196/38624
PMID:36301590

Review

Barriers to and Facilitators of Using eHealth to Support Gestational Diabetes Mellitus Self-management: Systematic Literature Review of Perceptions of Health Care Professionals and Women With Gestational Diabetes Mellitus

Ladan Safiee[1], MSc; Daniel John Rough[1], PhD; Heather Whitford[2], PhD

[1]School of Science and Engineering, University of Dundee, Dundee, United Kingdom

[2]School of Health Sciences, University of Dundee, Dundee, United Kingdom

Corresponding Author:
Daniel John Rough, PhD
School of Science and Engineering
University of Dundee
Queen Mother Building
Dundee, DD1 4HN
United Kingdom
Phone: 44 01382 383000
Email: drough001@dundee.ac.uk

Abstract

Background: Gestational diabetes mellitus (GDM) is one of the most common medical complications during pregnancy. eHealth technologies are proving to be successful in supporting the self-management of medical conditions. Digital technologies have the potential to improve GDM self-management.

Objective: The primary objective of this systematic literature review was to identify the views of health professionals (HPs) and women with GDM regarding the use of eHealth for GDM self-management. The secondary objective was to investigate the usability and user satisfaction levels when using these technologies.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach, the search included primary papers in English on the evaluation of technology to support self-management of GDM from January 2008 to September 2021 using MEDLINE, CINAHL, Embase, ACM, and IEEE databases. The lists of references from previous systematic literature reviews, which were related to technology and GDM, were also examined for primary studies. Papers with qualitative, quantitative, and mixed methodologies were included and evaluated. The selected papers were assessed for quality using the Cochrane Collaboration tool, National Institute for Health and Care Excellence clinical guidelines, Critical Appraisal Skills Programme Qualitative Checklist, and McGill University Mixed Methods Appraisal Tool. NVivo (QSR International) was used to extract qualitative data, which were subjected to thematic analysis. Narrative synthesis was used to analyze the quantitative data.

Results: A total of 26 papers were included in the review. Of these, 19% (5/26) of studies used quantitative research methodologies, 19% (5/26) used qualitative methods, and 62% (16/26) used mixed methods. In all, 4 themes were identified from the qualitative data: the benefits of using technology, engagement with people via technology, the usability of technology, and discouragement factors for the use of technology. The thematic analysis revealed a vast scope of challenges and facilitators in the use of GDM self-management systems. The challenges included usability aspects of the system, technical problems, data privacy, lack of emotional support, the accuracy of reported data, and adoption of the system by HPs. Convenience, improved GDM self-management, peer support, increased motivation, increased independence, and consistent monitoring were facilitators to use these technologies. Quantitative data showed that there is potential for improving the usability of the GDM self-management systems. It also showed that convenience, usefulness, increasing motivation for GDM self-management, helping with GDM self-management, and being monitored by HPs were facilitators to use the GDM self-management systems.

Conclusions: This novel systematic literature review shows that HPs and women with GDM encountered some challenges in using GDM self-management systems. The usability of GDM systems was the primary challenge derived from qualitative and

quantitative results, with convenience, consistent monitoring, and optimization of GDM self-management emerging as important facilitators.

(J Med Internet Res 2022;24(10):e39689) doi:10.2196/39689

KEYWORDS

gestational diabetes mellitus; GDM; gestational diabetes; self-management; eHealth

Introduction

Background

Gestational diabetes mellitus (GDM) is defined as any degree of carbohydrate intolerance with onset or first recognition during pregnancy [1]. GDM is one of the most common medical complications of pregnancy [2], with a significant increase in its prevalence in different ethnic groups and countries over the last several years [3,4]. GDM is most prevalent in the Middle East and North Africa, with an estimated median of 12.9%, and least prevalent in Europe, with an estimated median of 5.8% of all pregnancies [5]. In the United Kingdom, the prevalence of GDM is approximately 4% of all pregnancies [6]. The rate of GDM is likely to rise owing to a growth in GDM risk factors, such as greater prevalence of maternal obesity and advancing age of childbearing [7], leading to an increasing demand for GDM clinical services [8].

GDM is associated with serious maternal [9-11] and fetal complications [12-15]. Mothers who have been affected by GDM are also at risk of developing type 2 diabetes [16] and cardiometabolic disorders later in life [17], and their infants are more at risk of developing adulthood obesity and type 2 diabetes [12,18]. These complications represent significant health problems and cost [19] for health services. The risk of adverse effects of GDM can be minimized by good control over maternal blood glucose (BG), diet, and physical activities [20]. However, there is limited time between diagnosis and delivery to optimize care for women with GDM [21]. Therefore, regular clinic visits [22] to a multidisciplinary team are advised to provide care during pregnancy. Nonetheless, traveling to specialist clinics in central locations [23] is expensive [24], time consuming, and inconvenient for women [25]. Recently, there has been an increase in the use of technology to enable self-management of GDM by women and to shift GDM management away from hospital-based care [26].

In light of increased adoption of technology to access information and communication, a digital GDM self-management system might offer advantages such as reducing patient travel and waiting time [27], saving medical practitioner time [8], reducing costs [28,29] to both the health care system and patients, greater convenience [30], attainment of better pregnancy outcomes [31], and an increased feeling of self-efficacy [32]. This can further lead to better BG control [29,33] and a decrease in GDM complications owing to greater accuracy and more frequent monitoring [34]. Such outcomes are evident in the results of several studies, which have found that health care technology can be beneficial for women with GDM in the improvement of hemoglobin A_{1c} [35-37], mean BG [21,38-40], maternal weight [41], and maternal and fetal outcomes [38,42,43]. Technology could also offer high-quality remote health care in a critical situation such as the COVID-19 pandemic to women with GDM, where travel and in-person contact have been severely restricted [44,45]. Therefore, there is an urgent need to consider computer-based communication technologies for the management of diabetes. This could contribute to better diabetes management by improving patient knowledge, attitudes, skills, lifestyle behavior [46], quality of care, and access to care [29].

Study Aims

Digital GDM self-management systems developed in recent years are available mostly as mobile apps or websites [8,30,34] and offer a wide range of features such as monitoring BG [23], diet, physical activity, blood pressure, and ketonuria [8] for women with GDM. However, a recent study by Kalhori et al [47] suggests that the few GDM apps available in popular app stores are poor in quality, using the Mobile App Rating Scale as a basis for this result [47].

Furthermore, most GDM self-management systems are not widely used [48,49], and some are no longer supported [8,50], one reason for which is obsolete hardware (ML Bartholomew, MD, email communication, 2018). Previous systematic reviews in the scope of technology and GDM management were carried out on available technology for GDM self-management [47,51-53], the impact of technology on clinical and pregnancy outcomes or GDM management [54-56], comparing women's clinical outcomes using technology with standard care [35], and the psychological aspect of using technology [57]. However, to the best of our knowledge, there is no systematic literature review of the opinions of health care professionals and women with GDM about using technology for GDM self-management.

The primary aim of this systematic literature review was to identify the views of health professionals (HPs) and women with GDM regarding barriers and facilitators of using technology for GDM self-management. The secondary aim was to investigate the usability and user satisfaction of these technologies.

Methods

Approach

The search strategy was developed by following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach [58] with the help of a professional librarian. The PRISMA guidelines lead to standardized reports and enhance the clarity of systematic literature reviews [59].

Criteria of Inclusion and Exclusion

To achieve the aims of this review, the criteria for inclusion and exclusion were developed as presented in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

1. Views of health care professionals, pregnant women diagnosed with gestational diabetes mellitus (GDM) or postpartum women with a history of GDM about their pregnancy period
2. Technology (eHealth or telemedicine being used, evaluated, reviewed, or discussed by participants) or usability evaluation or reports of user satisfaction levels
3. Any primary research studies
4. Aspects of GDM management (eg, blood glucose control, diet, weight, physical activity, medication adherence, or information)

Exclusion criteria

1. Published papers written in any language other than English
2. Women with preexisting type 1 and type 2 diabetes (except papers that provide information about GDM distinct from type 1 and 2 diabetes)
3. Any nondigital technology
4. Papers published before 2008
5. Posters, abstracts, and news items
6. Systematic literature reviews
7. Usability results for task performance

Search Strategy and Screening Process

A search was carried out using 3 search terms—"self-management," "gestational diabetes," and "technology" (Multimedia Appendix 1). The search terms were identified from papers in eHealth for GDM in the PubMed database.

The search included publications written in English from January 2008 to September 2021 in the MEDLINE, CINAHL, Embase, ACM, and IEEE databases. This date limitation was chosen to represent contemporary technology for GDM self-management.

The screening process was conducted by the first author in line with previous studies [60,61] and with the help of the research team and a professional librarian using the following steps:

1. Identification: the results of the search from different databases were exported to the EndNote X7 software. Furthermore, the reference lists of previous systematic literature reviews related to technology and GDM were examined in the primary studies. All citations were collated into one group and duplicate records were removed.
2. Screening: the titles and abstracts of the remaining citations were screened based on the inclusion and exclusion criteria to select potential papers by the first author. At this stage, 2 other members of the research team independently conducted a double screening of the first 10% of the results. Following a discussion phase, this screening process was repeated to ensure reliability based on inclusion and exclusion criteria.
3. Eligibility: Mendeley software was used to keep electronic copies of the full text of potential papers. The full text of the papers was assessed based on the inclusion and exclusion criteria.
4. Included: the final papers were selected from the full text based on the inclusion and exclusion criteria by the first author. The papers were discussed with the research team if there was any lack of clarity in their inclusion.

Data Extraction

The study characteristics were extracted from the final 26 included papers. A predefined data extraction table was populated with information, such as study design, sample size, location, analysis method, participants' ages, inclusion and exclusion criteria, analysis methods, study goals, quantitative and qualitative data collection tools, and key findings (Multimedia Appendix 2 [13, 21, 25, 27, 30, 34, 41, 43, 48, 50, 62-77]).

NVivo 12 was used to extract relevant qualitative data to achieve the primary aim of the review. A predefined table, including the author, measures, scale items, and results, was used to extract relevant quantitative data.

Quality Assessment

Appropriate appraisal tools were chosen based on the methodology and study design. Each of the studies included in this review was critically assessed using an appropriate tool: the Cochrane Collaboration tool for randomized controlled trials (RCTs) [78], National Institute for Health and Care Excellence clinical guidelines for questionnaire studies or surveys [79], the Critical Appraisal Skills Programme Qualitative Checklist for qualitative studies [80], and the McGill University Appraisal Tool for Mixed Methods [81].

To meet the aims of this systematic literature review and not to exclude data relevant to this review, the quality of papers was not assessed with the purpose of excluding them. Instead,

limitations of the included studies were considered during the analysis and synthesis of data.

Analysis

The analysis was completed in 2 phases for qualitative and quantitative data. Thematic analysis with an inductive approach [82] was used to develop themes from 73% (19/26) studies that included qualitative data following the 6 steps outlined by Braun and Clarke [82].

Level 1 (reviewing codes of each theme for existence of coherent patterns) and level 2 analyses (reviewing the themes to assess whether they reflect the entire data set) were conducted by the first author and the second coauthor. Interrater reliability was not carried out, in line with the recommended process by Braun and Clarke [83].

Narrative review was used to analyze the quantitative data owing to the heterogeneity of research methods used. A narrative review is flexible and allows different types of evidence to be combined into a coherent summary. The narrative review process [84] included summarizing and explaining the quantitative data presented in 69% (18/26) included papers.

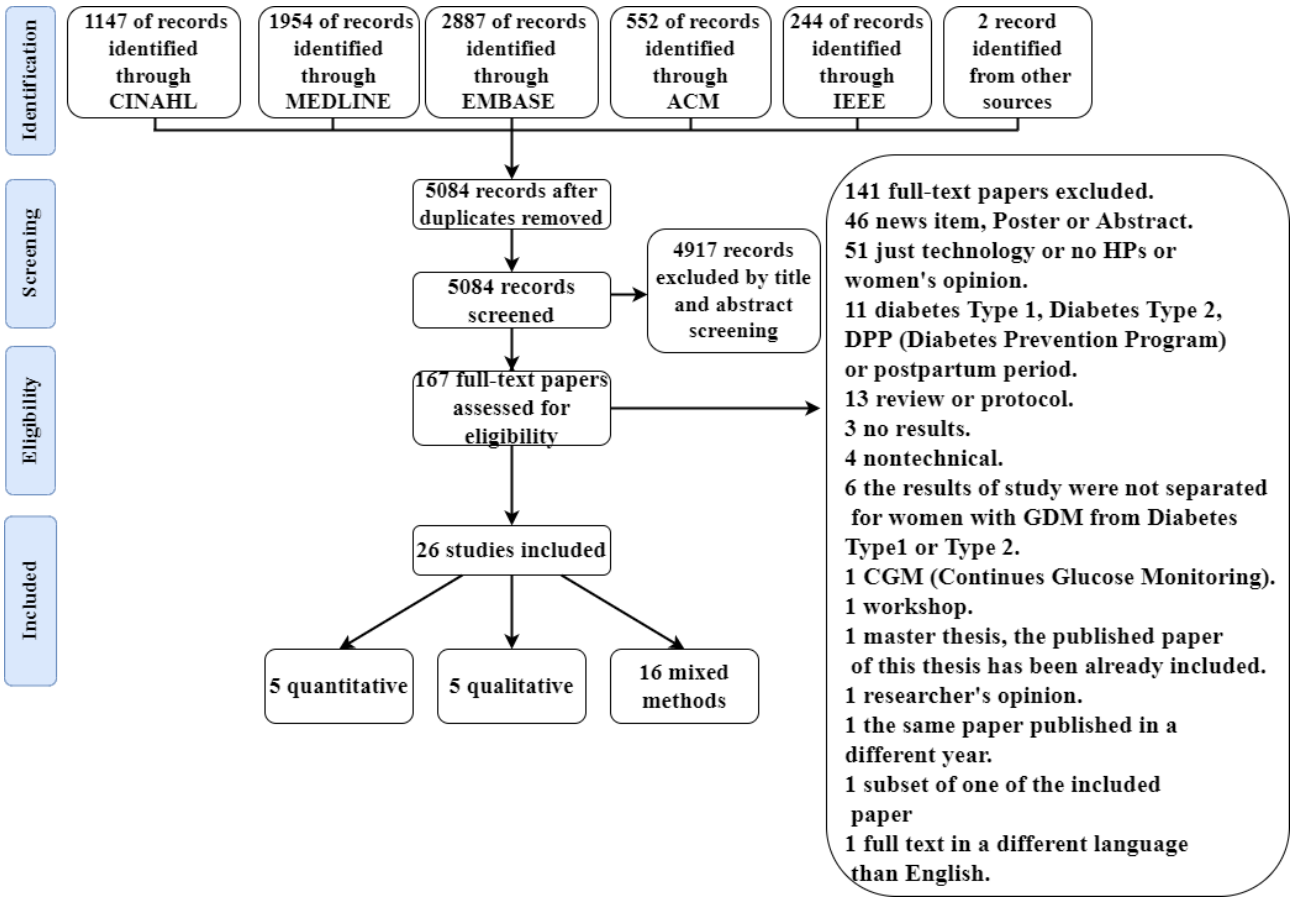
Results

Study Selection and Study Characteristics

The search and screening strategies are shown in Figure 1.

A total of 26 papers were included from the full text based on the inclusion and exclusion criteria. Of the included papers, 19% (5/26) were quantitative, 19% (5/26) were qualitative, and 62% (16/26) used mixed methods (Multimedia Appendix 2). The sample sizes varied among the studies, ranging from 9 [62] to 340 [63] participants. Most of the included studies were from Europe (15/26, 58%), and the rest were from North America (3/26, 11%), Australia (4/26, 15%), Singapore (1/26, 4%), New Zealand (1/26, 4%), and South Korea (1/26, 4%), with 4% (1/26) study of unspecified location. Studies varied in exploring the views of women and HPs. Of these, 96% (25/26) studies included the views of women, with 23% (6/26) including the views of HPs, and only 4% (1/26) including HPs' views without those of women.

Figure 1. Study identification flowchart. GDM: gestational diabetes mellitus; HP: health professional.



Methodological Quality Assessment

In general, the 26 included studies showed some degree of bias in their research.

Figure 2 [21,25,34,43,50,64] and Figure 3 show the risk of bias summary and graph (specific to an RCT study design), respectively, for the included studies using Review Manager 5.3 (Cochrane Collaboration desktop software).

On the basis of the nature of the included studies that used technology as a core of their research, it was impossible to blind participants and researchers from the knowledge of the intervention participants received [35]. Therefore, performance bias was not included in the risk of bias assessment (Figures 2

and 3) [35]. Of the 23% (6/26) RCT studies, quality assessment showed that 15% (4/26) had a low risk of bias [21,25,43,64]. The other 8% (2/26) studies presented a risk of bias in incomplete outcome data owing to the withdrawal of a large number of participants during the study [50] and an unequal number of participants in the intervention and control groups [34]. Furthermore, the allocation concealment method has been adequately reported in only 8% (2/26) studies [25,50].

Quality appraisal of the remaining studies (Multimedia Appendix 3 [13,21,25,27,30,34,41,43,48,50,62-77]) revealed that 11% (3/26) qualitative studies were of good quality in design, data collection procedure, and data analysis [62,65,66]. The common limitations for the rest of the studies (including quantitative, qualitative, or mixed methods) were bias in sampling [49,67,68], small sample sizes relative to the type of study conducted [67,69,70], lack of information about the validity and reliability of the data collection tools [67,69-71], lack of information about inclusion and exclusion criteria [68], poor qualitative results [64], and unclear recruitment strategy [72]. In addition, there was a lack of information regarding the method of gathering qualitative data [13] and the analysis process [13,72]. In 8% (2/26) mixed methods studies, it was stated that the quantitative data would be collected in the following phase, but there was no clear explanation about how the triangulation of the quantitative and qualitative data would answer the research question [27,48].

Figure 2. Risk of bias summary—each risk of bias item across included randomized controlled trial studies. Green: Yes (low risk of bias); Red: No (high risk of bias); Yellow: Unclear (bias is not clear or bias cannot be determined) [21,25,34,43,50,64].

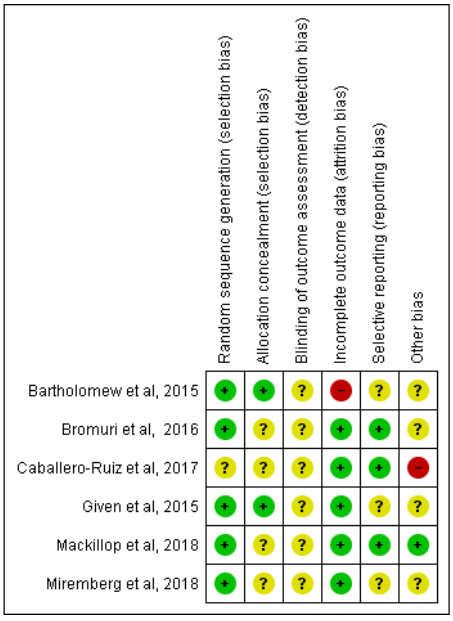
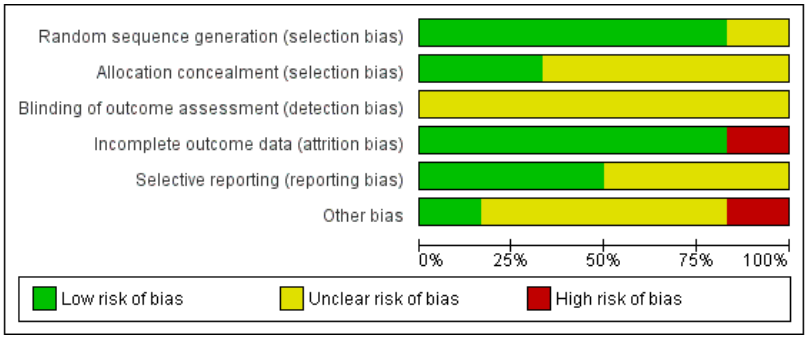


Figure 3. Risk of bias graph—the risk of bias item presented as percentages across included randomized controlled trial studies. Green: Yes (low risk of bias); Red: No (high risk of bias); Yellow=Unclear (bias is not clear or bias cannot be determined).

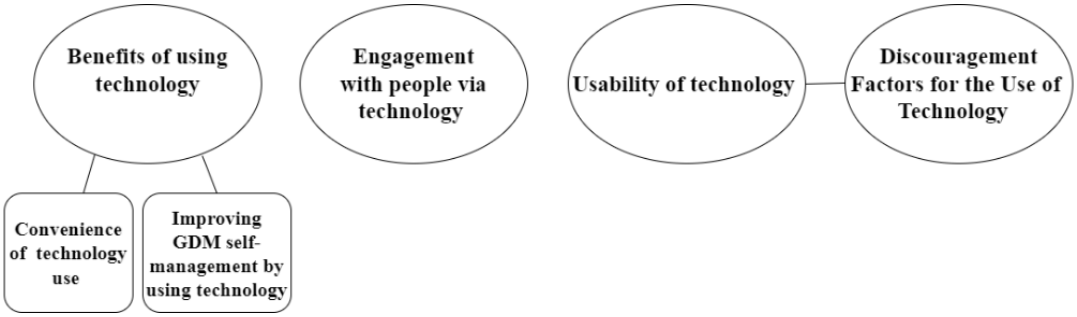


Thematic Analysis of Qualitative Data

Overview

Of the included studies, 73% (19/26) contributed qualitative data to the thematic analysis. The views of women and HPs were integrated and reported together throughout the analysis. A total of 4 themes were identified: benefits of using technology, engagement with people via technology, usability of technology, and discouragement factors for the use of technology (definitions of the themes and subthemes are available in Multimedia Appendix 4). Furthermore, 2 subthemes were identified, as outlined in Figure 4.

Figure 4. Thematic map showing themes and subthemes. GDM: gestational diabetes mellitus.



Theme 1: Benefits of Using Technology

Overview

Both women and HPs reported their confidence in [27,72] and willingness to use GDM self-management systems because of the benefits of these systems for women with GDM [25,30] and for their babies' health [65,66,72]. Furthermore, some HPs considered technology to be beneficial for complementing the limited number of health care professionals, while the rate of GDM is increasing [66]. The benefits of using technology themes included 2 subthemes: "convenience of technology use" and "improving self-management by using technology."

Convenience of Technology Use

Convenience was the predominant benefit of using technology for GDM management. A total of 50% (13/26) papers reported that women with GDM and HPs found the convenience of reduced travel and clinical appointments, as well as the pervasive use of technology, the most beneficial reasons for its use. Women in the studies of Khalil [66] and Edward et al [73] expressed that traveling is "exhausting" [66] particularly toward the end of their pregnancy [73], and especially for women living at a distance [27,30,66]. Women and HPs also indicated that it would lead to a reduction in the need for women to make potentially stressful arrangements for finding childcare and managing absence from work [25,27]. Therefore, technology could be highly advantageous for women with busy lives, especially those who already have children [30]:

> *I am amazed with the technology and it suited me much better than having to travel in a lot and wait, especially with little ones* [Patient 10]

Generally, women and HPs lauded the ease and convenience of using technology rather than traditional paper logbooks. This was mainly because of the ability to access technology anytime [74] or anywhere, driven by the growing pervasiveness of mobile devices [62,73]—"you've always got your phone haven't you, so it's the easiest way to do stuff" (Patient 3).

Women and HPs also recognized constant access to information related to GDM [62,65,73] and being familiar with using similar technology [48,62,72-74] as further elements of ease and convenience.

Saving time is another convenient aspect of technology use for both women [25,27,75] and HPs [64]. In a study by Bromuri et al [64], a telemedicine system helped HPs review BG values quicker than to review them on a paper logbook, owing to alerts that highlighted out-of-range BG values resulting in hyperglycemia and hypoglycemia being recognized quickly. Women and HPs agreed that it takes considerable time to attend clinical appointments [27] just to "be told you're doing everything right" [25,75]:

> *They don't want to spend all of their time trying to get to the hospital and look for parking and spend long periods waiting at hospital.* [Clinician 2]

Improving GDM Self-management by Using Technology

Improving the ability of women to self-manage GDM is another prominent benefit of using such technology. Increasing awareness of one's own data has been perceived as an important element of using technology for GDM self-management [25,63,65]. Women in some studies indicated that real-time feedback [63,65,74], visualization of data (eg, graphic nutrient summaries or recommendations) [63,65,74] and the ability to review and track their data [25,63,75] empowered them with "self-awareness" about their own data [63,65,74]. The clarity of the relationship between different attributes, particularly diet and BG levels, was seen as beneficial [63,65,73]. Data relationships also helped women to identify "where it [self-management] was working or where it was going wrong" [25] and supported them to change their lifestyle [63,65,74]. However, women and HPs had different opinions about the accuracy of women's self-reported data. Although some women favored recording data with technology because they were more accurate and precise [74], other women admitted misreporting their data values to get more positive feedback [65]. Some HPs did not want to rely on women's self-reported data [76] because they did not trust the accuracy of the data; they preferred to enter data into the system themselves [72].

Women also found information related to diet [49,63,68,73,74] and peer support [73,75] useful in improving their lifestyle. Moreover, women felt "automatic messages" [63], rewards, and goal tracking on the system motivated them to change their lifestyle and optimize their GDM self-management.

Both women and HPs perceived increased independence through technology [66,72]. Women and HPs also expressed that using a digital GDM system improved both their self-management skills [27,65,66] and exercise of control on their GDM condition [25,27,63,65,71,73,74,77]:

> *myDiabby helped patients self-manage their health. [66]* [Nurse 2]
>
> *Technologies help us being more autonomous. We feel more responsible. [66]* [Patient 1]

Theme 2: Engagement With People via Technology

This theme included 2 main components including engagement with peers and engagement with health care professionals.

Women with GDM indicated that accessing "peer support" by a digital GDM self-management system would be useful [75] as "somebody may know something more" [68]. Although some women had little or no experience with web-based group communication, they were still interested in communicating with other women with GDM via technology [68]. Peer support provided an opportunity for women to access "other people's experiences" [73] for sharing and exchanging information [68]. As a woman with GDM indicated, peer support provided "a better overview of risks associated with GDM, what could go wrong potentially, and the good stories as well" [73]. Overall, women perceived that peer support empowered them with a broader scope of GDM knowledge than other women who were experiencing the same condition [73]. In addition, peer support reassured women that they were "not alone" [73] and offered them a "constant feeling of support" [73]. Furthermore, it enabled women to talk about their condition and experience in a "safe space" without being judged by other people [73]. Women indicated a lack of peer support in the current care system that might be addressed using technology [75].

Regarding engagement with health care professionals, women appreciated the possibility of receiving additional support using technology. They valued sharing their data and having regular GDM monitoring by HPs via technology [25,66,68,72,73,76], specifically for benefiting both their own health and that of their baby [72]. Women expressed how sharing data with HPs was "*reassuring*" and gave them a "*safety net*" [25] owing to a feeling of being monitored more closely by the HPs [73]. Similarly, some HPs believed that sharing data would provide an opportunity to review and monitor the data frequently [64,66], detect any changes or problems at an early stage [25,48] and thereby allow the early application of treatment or interventions for women with GDM [25,64,66].

Although some women and HPs felt comfortable communicating via technology [30,66,71,73,75], others were concerned about a lack of physical and emotional support [27,62,66,75] and a poorer quality of conversation [25,27,66]:

> *I like the one to one contact so you can ask questions. [25]*

Nevertheless, women still felt there was a need to provide more interaction and communication between HPs and women via mobile app technology [76].

Theme 3: Usability of Technology

Women and HPs provided various perspectives on the usability of digital GDM systems in this theme. The content of the systems, including the quality of information and format and presentation of patients' data, was the main usability aspect discussed by women and HPs in the included papers.

When women and HPs found the GDM systems "easy to use" [25,30,66,72,74], "simple" [25,66], "intuitive" [66] and "straightforward" [25,74], these impressions were influenced by the presence of simple language and images [63] and the simplicity of information presentation, such as displaying all data on one screen [62,72,77].

When usability concerns arose, they were also related to the data format and layout. Women and HPs suggested improving the layout and format of the information by changing the size of images or the amount of text [49], using videos [27], improving the data summary presentation [74,77], changing the data format to be similar to that of a paper logbook [25,48,63,74], and distinguishing different degrees of BG severity [48]:

> *To look back and see is there a blood sugar previous to try and identified yourself which was the pre and which was the post [meal test]. [25]*

In the study by Pustozerov and Popova [76], HPs also indicated that improving the data format would help them review the data more easily.

Discussions on usability were also directed at the effectiveness of GDM apps in fulfilling the needs of women. Participants in different studies provided opinions about the lack of functionality in their GDM self-management systems. Some of their diverse suggestions included an option to scan barcodes of food [74], a time-alerting function for entering data [13], an educational or coaching feature [48], the ability to add a note to BG readings, and the ability to record the type of physical activity they have performed [65].

Women were also interested in having pop-up messages [65], informing them about any changes in their data [48], their condition [73], or any new activities in the forum [68] on the system:

> *To be able to review previous (entered) results and comments, to get an alert notice if results are out of the ideal range... [48]*

A final aspect of usability concerned the effectiveness of information content. Women and HPs found the GDM information in both older technology [73] as well as that introduced by the studies [63,65,77] to be insufficient and simplistic [63,65,73,77]. Personalized information was considered vital [63] for diet [48,49] and in-depth information about GDM [65,77]. In addition, some women had issues with the clarity of the content and wanted simple, clear [49,77] and commonly used language [63,77] such as using "*tablespoon*" or "*bowl*" as familiar measurement units used by their dietitian rather than imperial measurements that were used to display food quantities on the app [63].

Despite the clear views of some women and HPs that using GDM self-management technology was more efficient in monitoring [48] and recording [68], other women were concerned about the inefficiency of their GDM systems [73,74]. Women with GDM found it was time consuming to use the system, particularly to retrieve information from food databases [74]. Postpartum women who had GDM perceived that the apps they used for GDM were overcomplicated and required too much commitment to complete a task [73]:

> *For something that was quite simple, it would take actually a long time to find it. [74]*

> *I've never managed to do it for a long period, because of the amount of commitment. [65]*

Theme 4: Discouragement Factors for the Use of Technology

The apparent disinterest of HPs was a cause of discouragement for women with GDM. Some said HPs lacked interest [63,65,71] and knowledge [65] in using technology. Indeed, their HPs' preference for a paper logbook discouraged women from using digital GDM self-management tools [65], particularly those who were already unfamiliar with such technology [27]:

> *I had no interest in writing it two places, and I understood that no one was going to read or use my app…They always asked for my book, so I used that. [79]*

Similarly, HPs were concerned about women's abilities to use technology:

> *they're all on their screens but at the end of the day, some of them don't actually have credit to even look at a website or download a piece of information. [67]*

Confirming this, some women reported little or no experience of using "message boards and things of that nature" [68]. Therefore, women themselves believed that some training might be needed to increase their confidence to use such technologies [27,72]. Some women with GDM were also concerned about the privacy of personal health information recorded on the systems [25,77].

In addition, HPs were concerned about the time required to use the systems and thought it would increase their workload [25,48,66,75]. They were also concerned that some women might not be able to afford the technology [75]:

> *We have some women who have got quite a low socioeconomic status, most of them still have phones...but not all have [mobile] data. [67]*

Women with GDM and HPs also experienced technical problems as barriers to the use of GDM self-management technology. Both women and HPs reported some difficulties with data transmission [62,65,71], problems with accessing technology [75], and poor access to the local internet [25,30].

Narrative Review of Quantitative Data

A narrative review was used to analyze the quantitative data, including the usability and user satisfaction results from 50% (13/26) of the included studies. Quantitative data from the remaining 27% (7/26) studies were not included in the analysis, as the results were not related to usability or user satisfaction [27,48,49,63,72,75] or were the result of objective task performance [77].

Usability

Quantitative studies used various measurements to gather data. Of these, only 12% (3/26) included a usability questionnaire to evaluate their systems, as summarized in Table 1.

Of the included studies, 8% (2/26) applied the system usability scale (SUS) developed by Brook in 1996 [85], with defined acceptability ranges for SUS scores (0-50 not acceptable, 50-70 marginal and 70-100 acceptable range) [86]. Jo and Park [13] reported a marginal score for their app, just below the acceptable threshold, (69.5 of 100). A similar, but acceptable, score was reported in the study by Gianfrancesco et al [74] for their web-based dietary system (70.9 out of 100) [74]. Pustozerov and Popova [76] included a custom questionnaire wherein women with GDM rated the "usefulness" and "convenience" of their GDM system on a 10-point scale. Usefulness was rated highly (8.7 out of 10), with convenience scoring somewhat lower (7.2 of 10).

In short, although these results suggest that previous GDM systems have usability challenges, it is impossible to draw any reliable conclusions with only 12% (3/26) studies providing results from a usability questionnaire.

Table 1. Included studies that used a usability questionnaire.

Usability	Type
Gianfrancesco et al [74]	SUS[a] questionnaire
Jo and Park [13]	SUS (Korean version)
Pustozerov and Popova [76]	Custom usability questionnaire (10-point scale questions on *convenience* and *usefulness* + open-ended questions)

[a]SUS: system usability scale.

User Satisfaction

The included studies used different measurements to evaluate user satisfaction. Given et al [25] used an adapted version of the Telemedicine Satisfaction and Usefulness Questionnaire by Bakken et al [87]. Of the studies that included user satisfaction questionnaires, 4% (1/26) did not make their satisfaction questionnaire available [21], 12% (3/26) used specially developed satisfaction questionnaires [30,34,50], and the rest (4/26, 15%) used satisfaction questionnaires without any information on how they were developed [21,67,69-71]. Studies by Hirst et al [30] and Mackillop et al [43] were the only ones to provide evidence of the validity and reliability of their developed questionnaires.

The included studies reported generally high user satisfaction in their evaluations of GDM systems [21,25,30,34,50,67,69-71]. However, their user satisfaction questionnaires evaluated many different aspects of GDM systems regarding the type of technology and its features, making it difficult to clearly summarize areas for improvement. Table 2 shows the key measures of the user satisfaction questionnaires in the included studies (the complete measures are available in Multimedia Appendix 5). Most questionnaires used a Likert scale rating to assess the degree of participants' agreement with their

statements about the GDM systems. Women in these studies interacted with the technology within the period from GDM diagnosis until childbirth (usually between 8 and 10 weeks). They all used and evaluated the real working prototypes. Miremberg et al [21] were not included in Table 2 because the questions or satisfaction items were not available in their study.

Assessment of the aspects of convenience was common. Caballero-Ruiz et al [34] highlighted the convenience of minimizing travel to centralized clinics as the strongest indicator of satisfaction (approximately, on average, 9.5 out of 10). In other studies, women rated GDM apps highly for factors such as not complicating their lives [34,69,70] and the ability of these apps to fit into their lifestyles [30].

Improvement of GDM self-management was a highly rated aspect of the studied systems, including helping women to record BG levels [71], reminding them to take medication and record BG levels, helping them eat healthier, encouraging them to be more active [67], and helping to improve their GDM knowledge [34]. Moreover, most women found SMS text messages helpful and motivated them to optimize their GDM self-management [50,67]. A total of 2 studies also reported a general increase in women's confidence in the management of their GDM [34,70].

Confidence or trust in GDM systems was rated well. Women with GDM reported confidence that the health care team checked their BG levels on the GDM system [71]. Many studies reported high ratings of confidence in the GDM systems, with women recommending them to others [34,43,50,67,69] or planning to use them in their next pregnancy [25,43,50,67]. Similarly, a study reported a high degree of trust (average 9 out of 10) in the GDM system [34], while another study reported that the GDM system was reliable [30].

Slightly lower satisfaction scores were reported for other aspects of ease of use: clarity of visualization of changes to treatment was rated approximately 7 out of 10 [34], and Peleg et al [70] reported satisfaction with system response time as approximately 3.5 out of 5, and ability to assist with interpreting self-monitored data approximately 3.8 out of 5.

Overall, based on the usability results (scores just under or above the acceptable threshold), there is much room for improvement in the usability of GDM self-management systems. However, with the limited number of papers providing a quantitative usability evaluation and the heterogeneity of questions assessing satisfaction, more studies are needed to identify where the improvement of usability and user satisfaction should be focused.

Table 2. User satisfaction question topics in the included studies.

Summary of key measures of user satisfaction questionnaires	Study								
	Varnfield et al [71]	Johnson and Berry [67]	Mackillop et al [43]	Peleg et al [70]	Caballero-Ruiz et al [34]	Peleg et al [69]	Bartholomew et al [50]	Hirst et al [30]	Given et al [25]
Convenient			✓[a]					✓	
Avoiding displacement					✓				
Fit in with life or did not complicate it			✓	✓	✓	✓		✓	
Adapt to daily life and context changes				✓		✓			
Number of hospital consultations is enough			✓		✓				
Help to record BGLs[b]	✓								
Help to remember to take medication and take BG[c]		✓							
Help to eat healthier or become more active		✓							
Helps to improve GDM[d] knowledge					✓				
Increased motivation for self-management							✓		
Improved diabetes control							✓		
Help to feel confident in managing GDM	✓			✓		✓			
Feel confident that health care team checked BGLs	✓								
Recommending to others		✓	✓	✓	✓	✓	✓		
Using it again		✓	✓	✓		✓			✓
Useful				✓	✓				
Easy to use				✓			✓		✓
Ease to learn how to use				✓	✓				
Helps data interpretation				✓	✓				
Clarity or effectiveness of visualization				✓	✓				
Clarity of activities' sequence in app				✓					
Personalized							✓		
System response time				✓		✓			
Experiencing error with the system				✓					
Time consuming							✓		
Trust is being well controlled					✓				
Trust it to work									✓
Reliable to use			✓					✓	
Satisfaction regarding diabetes follow-up					✓				
Satisfied with the system	✓								✓
Enjoyable or interesting				✓			✓		
Paying for the system				✓		✓			

[a]✓: illustrates where a study included a measure of user satisfaction in its participant questionnaire.

[b]BGL: blood glucose level.

[c]BG: blood glucose.

[d]GDM: gestational diabetes mellitus.

Discussion

Principal Findings

Overview

The primary objective of this systematic literature review was to identify the views of HPs, women with GDM, and postpartum women who have had GDM regarding GDM self-management technology. The secondary objective was to investigate the usability and user satisfaction levels of existing technologies and quantitatively evaluate these factors.

Regarding the first objective, thematic analysis of the qualitative data in the selected papers identified four themes: (1) the benefits of using technology, (2) engagement with people via technology, (3) usability of technology, and (4) discouragement factors for the use of technology.

The thematic analysis of qualitative data revealed barriers to usability, including technical problems, data privacy, lack of emotional support, the accuracy of reported data, and adoption of the system by HPs. Convenience, improving GDM self-management, peer support, increasing motivation, increasing independency, and providing consistent monitoring were common facilitators of using this technology.

For the second objective, the narrative review of the quantitative data (usability and user satisfaction) showed that there is room for improvement in the usability of GDM self-management systems.

Benefits of Using Technology

Convenience of Technology Use

The influence of convenience in our analysis, in both the qualitative and quantitative findings, is echoed in other literature on telemedicine. Pérez-Ferre et al [88] reported a 65% reduction in the number of clinical visits for women with GDM who were using telemedicine. The main benefits of doing so are the improvement of HPs' work efficiency and a better quality of life for women with GDM [57].

Although our findings indicated a strong positive desire to reduce in-person clinics through technology, not everyone wanted clinical visits replaced altogether. This was affirmed in a recent systematic review that highlighted the negative impact of losing in-person contact between women with GDM and HPs [57], particularly for women who experience social isolation and anxiety during pregnancy [89]. However, these studies were carried out before the COVID-19 pandemic. Today, patients may be more familiar with remote consultations, and the impact of this would benefit from further investigation.

Improving GDM Self-management by Using Technology

Our results revealed that women appreciated the use of technology to manage various aspects of their condition. These findings are consistent with those of relevant studies outside the scope of this review. Leziak et al [90] explored the experiences of women with GDM and pregnant women with type 2 diabetes using mobile health (mHealth) during pregnancy. Their results showed enhanced self-management through the use of mHealth technology [90]. Similarly, Yee et al [91] explored how pregnant women with GDM or preexisting diabetes perceived an SMS-based intervention during their pregnancy, showing an optimization of GDM self-management and increased motivation for diabetes self-care. In 2007, Homok et al [32] evaluated the feasibility of a web-based telemedicine system that monitored the BG levels of underserved (poor socioeconomic status) women with GDM using the Diabetes Empowerment Scale [92]. Participants experienced increased diabetes management self-efficacy, such as readiness to change their lifestyle behaviors to achieve diabetes goals.

In summary, evidence suggests that technology could help women optimize their GDM self-management abilities, leading to benefits for both themselves and their baby's health. As a result of good practices initiated through GDM self-management technology, women could also improve control over their health, which could be maintained habitually after giving birth to prevent the development of type 2 diabetes.

Engagement With People via Technology

As mentioned earlier, this theme consists of 2 main components: "engagement with peers" and "engagement with health care professionals."

The results of the thematic analysis demonstrated the benefits of peer support in digital GDM self-management systems [68,73,75] a finding supported by similar studies outside the scope of this review. Leziak et al [90] explored the experiences of low-income women with GDM and pregnant women with type 2 diabetes, using mHealth technology to support and improve diabetes self-management during pregnancy. Their results highlighted how women valued social interactions with other women and accessed their knowledge and experiences. McMillan et al [93] evaluated mHealth technology to support postpartum women with a history of GDM in maintaining postnatal activity and good dietary habits, finding that a discussion forum was a valuable feature in doing so [93]. As other previous studies have emphasized, such favorable opinions of women toward peer support stem from their ability to share or read stories about other women [91] and receive emotional support [94], which is an important factor in health communication [95,96]. Indeed, some HPs believed that pregnant women valued other women's experiences more than HPs' advice during their pregnancy [97]. However, Sherman and Greenfield [94] found that, when examining message boards for pregnant teenagers, some of the medical information posted by pregnant women was misleading because it was suitable for their specific condition and therefore inappropriate for others [94]. Furthermore, validation of posted information is also a major challenge [95], and further work is needed in this area to

provide a reliable and validated communication path between women with GDM.

Our thematic analysis described women's interest in sharing data with their clinicians by remote means, to obtain reassurance and to be monitored more consistently. This is also evident in some previous studies. Dalfra et al [31] found that women with GDM and pregnant women with type 1 diabetes appreciated their telemedicine system for sharing their data with HPs and their ability to communicate with them whenever needed. Similarly, Leziak et al [90] showed that women were also in favor of sharing data with HPs and receiving real-time feedback. However, in the included studies, some HPs found it difficult to trust women's reported data [72,76]. In contrast, Kruger et al [98] found that HPs were satisfied with the accuracy of the data reported by women with GDM via a telemedicine system. Other studies have found that it is unlikely that women would misreport their records, as they are highly motivated to maintain BG control [31] for the sake of their baby's health [57]. Further work is needed to examine the means of decreasing the possibility of reporting incorrect data.

Usability of Technology

Although the evidence available regarding the usability of digital GDM self-management systems is limited [99], the findings of our review are in line with those of previous studies on mHealth self-management systems for type 1 and type 2 diabetes. Katz et al [100] assessed 8 current diabetes self-management apps for adults with type 1 diabetes, discovering issues in the interpretability of data and high cognitive load. These results were corroborated by Fu et al [101] in an evaluation of 4 apps for type 2 diabetes management. Further studies have also found usability challenges with data format on mHealth self-management systems [102-104], such as difficulty interpreting or understanding data in its current format [104]. A useful digital self-management system should display data trends and patterns, specifically showing which data are normal or abnormal. Usability issues with data formats thus prevent patients from understanding their data [105,106], thereby limiting their self-management capabilities.

Our review also identified limitations in the functionality of the systems as another usability concern across the included studies. Previous reviews of general diabetes self-management apps have highlighted important missing functionality, including automatic transfer of BG data from a glucometer to a mobile app, personalized diabetes management advice [107], prevention of errors [108], freedom to edit or remove data entries and appointments, and the ability to automate common tasks [109].

The limited functionality of diabetes self-management systems can be considered a usability problem [109] and is likely to result in these systems failing to meet users' needs [107]. Addressing these functionality limitations would mitigate some of the usability challenges and help users optimize their engagement and interaction with these systems.

Quantitative evaluation of GDM self-management apps in the studies by Jo and Park [13] and Gianfrancesco et al [74] yielded SUS scores just below and above the acceptable threshold, respectively. Unsurprisingly, previous studies that used the SUS questionnaire to evaluate diabetes self-management apps in different domains have received similarly poor ratings [101,110,111]. Similar to this systematic review, these previous studies used guidance from Bangor et al [85] to interpret the SUS scores, with most apps falling below the acceptable range.

Our quantitative analysis identified the need to improve the usability of GDM self-management systems. However, with the limited number of papers providing a quantitative usability evaluation, the heterogeneity of questions assessing satisfaction and the variation in systems being assessed, it is difficult for quantitative studies to identify where the improvement of usability and user satisfaction should be focused. Therefore, it is an aspect that needs further investigation.

Discouragement Factors for the Use of Technology

Despite the perceived benefits of GDM technology, our analysis revealed technical problems as a prevalent barrier across the included studies. Previous studies have reported similar technical problems when using eHealth and self-management systems [102,112-115]. Moreover, a previous systematic literature review by Simblett et al [116] identified technical problems as one of the most significant barriers to using mHealth technologies. The most common technical problems in their review were app disappearance, loss of power, restarting without warning, not receiving notifications, receiving them at the wrong time, and having a difficult connection. Indeed, 2 participants withdrew from one of the included studies because of difficulties with internet connectivity. Parallel to the findings of this review, technical problems were the cause of reducing participants' motivations [112,113] and even the cause of leaving the study by participants with other health conditions [114,116].

In addition to technical problems, the privacy of personal health information was a concern for some women. Simblett et al [116] also reported privacy concerns in one of the included studies. Although the use of advanced encryption algorithms and pseudoanonymization of personal data should address security and privacy challenges at the system level, it is important for future GDM systems to effectively communicate good security practices to reassure new users [117].

Although most women across all studies were interested in using self-management technology, some suggested that their HPs were disinterested. Similarly, Wake et al [118] recognized the lack of awareness and adoption of technology by HPs as an important barrier to using eHealth for diabetes self-management [118]. HPs' difficulty to accept technology was experienced in previous studies [119-121], influenced by difficulty integrating it with their workflow [102,121], lack of integration with the medical record system [120], or a lack of technical knowledge [116]. Further work is required to involve HPs in the design and development of GDM technology more effectively to reduce this barrier.

Limitations and Further Work

The strengths of this review were its application of a rigorous process in paper selection and summarizing results that include both qualitative and quantitative data to cover a wide scope of understanding. Although this systematic literature review was conducted by the first author, we mitigated the potential for bias

through a double screening of a proportion of papers' citations (title and abstract) by the entire research team, in line with previous systematic literature reviews published in JMIR. Two of the authors were also involved in theme development and the methods and results were reviewed by all authors.

Thematic analysis was restricted to the qualitative data contained in the papers (19/26, 73%). It is possible that the authors of the included studies did not report significant results. However, it is unlikely that the key findings were not reported in the original papers.

The details of the methods and methodologies applied were limited in some studies. The available evidence is also limited by several factors. First, some studies used small sample sizes. Methodologically robust trials of greater sizes are needed to confirm the findings of our review. Second, the number of quantitative studies that measured usability was limited. Third, most of the evaluations of satisfaction did not address the validity and reliability of the satisfaction questionnaires. Furthermore, some questions in the satisfaction questionnaires were generic. Using standard evaluation tools and valid questionnaires would offer consistent and robust results across different studies.

Overall, further work is required to improve the usability of GDM self-management systems. There is a need to evaluate the systems using various usability approaches [109,122,123] and larger samples to obtain broader usability perceptions and identify problems with the systems. Furthermore, more engaging elements in a GDM self-management system are needed to develop better emotional support for women. Work is needed to improve peer communication to develop more support for women with GDM.

Further work is also needed to assess the design and development process of these GDM self-management technologies that might help identify the source of these usability challenges.

Conclusions

This is the first systematic literature review to carry out a comprehensive review of the perspectives of HPs, women with GDM, and postpartum women who have had GDM about using technology for GDM self-management during pregnancy. Despite the existence of several studies on technology and GDM, information about the perceptions of women with GDM and HPs regarding GDM self-management technology is limited. More rigorous studies are needed to reveal evidence-based barriers to and facilitators of using existing GDM self-management systems.

Acknowledgments

The authors would like to take this opportunity to thank the Engineering and Physical Sciences Research Council (EPSRC) for the funding support and opportunity to conduct this research project. They also thank Professor Annalu Waller, Dr Rachel Menzies, and Mr Scott McGregor for providing their help and advice during this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords and search strategy.
[DOCX File , 13 KB - jmir_v24i10e39689_app1.docx]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 59 KB - jmir_v24i10e39689_app2.docx]

Multimedia Appendix 3
Quality assessment.
[DOCX File , 45 KB - jmir_v24i10e39689_app3.docx]

Multimedia Appendix 4
Definition of the themes and subthemes.
[DOCX File , 15 KB - jmir_v24i10e39689_app4.docx]

Multimedia Appendix 5
Satisfaction measurements.
[DOCX File , 20 KB - jmir_v24i10e39689_app5.docx]

References

1. World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. World Health Organization. 1999. URL: https://apps.who.int/iris/handle/10665/66040 [accessed 2022-10-21]
2. Mastrogiannis DS, Igwe E, Homko CJ. The role of telemedicine in the management of the pregnancy complicated by diabetes. Curr Diab Rep 2013 Feb;13(1):1-5. [doi: 10.1007/s11892-012-0352-x] [Medline: 23242646]
3. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. Diabetes Care 2007 Jul;30 Suppl 2:S141-S146. [doi: 10.2337/dc07-s206] [Medline: 17596462]
4. Anna V, van der Ploeg HP, Cheung NW, Huxley RR, Bauman AE. Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. Diabetes Care 2008 Dec;31(12):2288-2293 [FREE Full text] [doi: 10.2337/dc08-1038] [Medline: 18809630]
5. Zhu Y, Zhang C. Prevalence of gestational diabetes and risk of progression to type 2 diabetes: a global perspective. Curr Diab Rep 2016 Jan;16(1):7 [FREE Full text] [doi: 10.1007/s11892-015-0699-x] [Medline: 26742932]
6. National Collaborating Centre for Women's and Children's Health. Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. In: The National Institute for Health and Clinical Excellence. London, UK: RCOG Press; 2008.
7. Births in Scottish Hospitals. Information Services Division Publication Report. 2017 Nov 28. URL: https://www.isdscotland.org/Health-Topics/Maternity-and-Births/Publications/2017-11-28/2017-11-28-Births-Report.pdf [accessed 2019-06-06]
8. Rigla M, Martínez-Sarriegui I, García-Sáez G, Pons B, Hernando ME. Gestational diabetes management using smart mobile telemedicine. J Diabetes Sci Technol 2018 Mar;12(2):260-264 [FREE Full text] [doi: 10.1177/1932296817704442] [Medline: 28420257]
9. Ovesen PG, Jensen DM, Damm P, Rasmussen S, Kesmodel US. Maternal and neonatal outcomes in pregnancies complicated by gestational diabetes. A nation-wide study. J Matern Fetal Neonatal Med 2015;28(14):1720-1724. [doi: 10.3109/14767058.2014.966677] [Medline: 25228278]
10. Pettitt DJ, Knowler WC, Baird HR, Bennett PH. Gestational diabetes: infant and maternal complications of pregnancy in relation to third-trimester glucose tolerance in the Pima Indians. Diabetes Care 1980;3(3):458-464. [doi: 10.2337/diacare.3.3.458] [Medline: 7389563]
11. Xiong X, Saunders LD, Wang FL, Demianczuk NN. Gestational diabetes mellitus: prevalence, risk factors, maternal and infant outcomes. Int J Gynaecol Obstet 2001 Dec;75(3):221-228. [doi: 10.1016/s0020-7292(01)00496-9] [Medline: 11728481]
12. Carolan M. Women's experiences of gestational diabetes self-management: a qualitative study. Midwifery 2013 Jun;29(6):637-645. [doi: 10.1016/j.midw.2012.05.013] [Medline: 22877761]
13. Jo S, Park H. Development and evaluation of a smartphone application for managing gestational diabetes mellitus. Healthc Inform Res 2016 Jan;22(1):11-21 [FREE Full text] [doi: 10.4258/hir.2016.22.1.11] [Medline: 26893946]
14. Metzger BE, Buchanan TA, Coustan DR, de Leiva A, Dunger DB, Hadden DR, et al. Summary and recommendations of the Fifth International Workshop-Conference on Gestational Diabetes Mellitus. Diabetes Care 2007 Jul;30 Suppl 2:S251-S260. [doi: 10.2337/dc07-s225] [Medline: 17596481]
15. Schneider S, Hoeft B, Freerksen N, Fischer B, Roehrig S, Yamamoto S, et al. Neonatal complications and risk factors among women with gestational diabetes mellitus. Acta Obstet Gynecol Scand 2011 Mar;90(3):231-237 [FREE Full text] [doi: 10.1111/j.1600-0412.2010.01040.x] [Medline: 21306307]
16. Bellamy L, Casas J, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
17. Krishnaveni GV, Hill JC, Veena SR, Geetha S, Jayakumar MN, Karat CL, et al. Gestational diabetes and the incidence of diabetes in the 5 years following the index pregnancy in South Indian women. Diabetes Res Clin Pract 2007 Dec;78(3):398-404 [FREE Full text] [doi: 10.1016/j.diabres.2007.06.002] [Medline: 17640759]
18. American Diabetes Association. 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]
19. Ohno MS, Sparks TN, Cheng YW, Caughey AB. Treating mild gestational diabetes mellitus: a cost-effectiveness analysis. Am J Obstet Gynecol 2011 Sep;205(3):282.e1-282.e7 [FREE Full text] [doi: 10.1016/j.ajog.2011.06.051] [Medline: 22071065]
20. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS, Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005 Jun 16;352(24):2477-2486. [doi: 10.1056/NEJMoa042973] [Medline: 15951574]
21. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
22. Sheikhi H, Saadatifar B, Dashtban R, Anvari N, Mastaelizadeh H. Self-care in patients with gestational diabetes. J Biochem Technol 2018;9(2):177-180.

23. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: 10.1177/1932296814542271] [Medline: 25004915]
24. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011 Jul;93(1):1-9. [doi: 10.1016/j.diabres.2011.02.002] [Medline: 21382643]
25. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
26. van den Heuvel JF, Groenhof K, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
27. Harrison TN, Sacks DA, Parry C, Macias M, Ling Grant DS, Lawrence JM. Acceptability of virtual prenatal visits for women with gestational diabetes. Womens Health Issues 2017;27(3):351-355. [doi: 10.1016/j.whi.2016.12.009] [Medline: 28153743]
28. Edwards L, Connors C, Whitbread C, Brown A, Oats J, Maple-Brown L, NT Diabetes in Pregnancy Partnership. Improving health service delivery for women with diabetes in pregnancy in remote Australia: survey of care in the Northern Territory Diabetes in Pregnancy Partnership. Aust N Z J Obstet Gynaecol 2014 Dec;54(6):534-540. [doi: 10.1111/ajo.12246] [Medline: 25308373]
29. Franc S, Daoudi A, Mounier S, Boucherie B, Dardari D, Laroye H, et al. Telemedicine and diabetes: achievements and prospects. Diabetes Metab 2011 Dec;37(6):463-476. [doi: 10.1016/j.diabet.2011.06.006] [Medline: 21889388]
30. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-115 [FREE Full text] [doi: 10.1177/1932296814556506] [Medline: 25361643]
31. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. J Telemed Telecare 2009;15(5):238-242. [doi: 10.1258/jtt.2009.081213] [Medline: 19590029]
32. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
33. Wojcicki JM, Ladyzynski P, Krzymien J, Jozwicka E, Blachowicz J, Janczewska E, et al. What we can really expect from telemedicine in intensive diabetes treatment: results from 3-year study on type 1 pregnant diabetic women. Diabetes Technol Ther 2001;3(4):581-589. [doi: 10.1089/15209150152811207] [Medline: 11911170]
34. Caballero-Ruiz E, García-Sáez G, Rigla M, Villaplana M, Pons B, Hernando ME. A web-based clinical decision support system for gestational diabetes: automatic diet prescription and detection of insulin needs. Int J Med Inform 2017 Jun;102:35-49. [doi: 10.1016/j.ijmedinf.2017.02.014] [Medline: 28495347]
35. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
36. Kim Y, Kim H, Kim Y. Effects of a web-based self-management program on the behavior and blood glucose levels of women with gestational diabetes mellitus. Telemed J E Health 2019 May;25(5):407-414. [doi: 10.1089/tmj.2017.0332] [Medline: 30036165]
37. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
38. Yew TW, Chi C, Chan S, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
39. Li S, Ouyang Y, Qiao J, Shen Q. Technology-supported lifestyle interventions to improve maternal-fetal outcomes in women with gestational diabetes mellitus: a meta-analysis. Midwifery 2020 Jun;85:102689. [doi: 10.1016/j.midw.2020.102689] [Medline: 32193015]
40. Seo Y, Kim EM, Choi JS, Park C. Using a mobile-based nutritional intervention application improves glycemic control but reduces the intake of some nutrients in patients with gestational diabetes mellitus: a case series study. Clin Nutr Res 2020 Jan;9(1):73-79 [FREE Full text] [doi: 10.7762/cnr.2020.9.1.73] [Medline: 32095450]
41. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019 Jan;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
42. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 06;20(1):198 [FREE Full text] [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]

43. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
44. Albert L, Capel I, García-Sáez G, Martín-Redondo P, Hernando ME, Rigla M. Managing gestational diabetes mellitus using a smartphone application with artificial intelligence (SineDie) during the COVID-19 pandemic: much more than just telemedicine. Diabetes Res Clin Pract 2020 Nov;169:108396 [FREE Full text] [doi: 10.1016/j.diabres.2020.108396] [Medline: 32890548]
45. Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]
46. Jackson CL, Bolen S, Brancati FL, Batts-Turner ML, Gary TL. A systematic review of interactive computer-assisted technology in diabetes care. Interactive information technology in diabetes care. J Gen Intern Med 2006 Feb;21(2):105-110 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00310.x] [Medline: 16390512]
47. Kalhori SR, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of english mobile applications for gestational diabetes: app review using mobile application rating scale (MARS). Curr Diabetes Rev 2021;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
48. Wickramasinghe N, Gururajan R. Innovation practice using pervasive mobile technology solutions to improve population health management: a pilot study of gestational diabetes patient care in Australia. J Healthc Qual 2016;38(2):93-105. [doi: 10.1097/JHQ.0000000000000033] [Medline: 26918811]
49. Carolan-Olah M, Steele C, Krenzin G. Development and initial testing of a GDM information website for multi-ethnic women with GDM. BMC Pregnancy Childbirth 2015 Jul 05;15:145 [FREE Full text] [doi: 10.1186/s12884-015-0578-0] [Medline: 26142482]
50. Bartholomew ML, Soules K, Church K, Shaha S, Burlingame J, Graham G, et al. Managing diabetes in pregnancy using cell phone/internet technology. Clin Diabetes 2015 Oct;33(4):169-174 [FREE Full text] [doi: 10.2337/diaclin.33.4.169] [Medline: 26487790]
51. Balaji B, Halperin I, Mukerji G, Lipscombe L. 1350-P: eHealth technologies for gestational diabetes mellitus: summary of features and effectiveness: scoping review. Diabetes 2020 Jun 1;69(Supplement_1):1350-P. [doi: 10.2337/db20-1350-P]
52. Garg N, Arunan SK, Arora S, Kaur K. Application of mobile technology for disease and treatment monitoring of gestational diabetes mellitus among pregnant women: a systematic review. J Diabetes Sci Technol 2022 Mar;16(2):491-497 [FREE Full text] [doi: 10.1177/1932296820965577] [Medline: 33118397]
53. Nguyen M, Hossain N, Tangri R, Shah J, Agarwal P, Thompson-Hutchison F, et al. Systematic evaluation of Canadian diabetes smartphone applications for people with type 1, type 2 and gestational diabetes. Can J Diabetes 2021 Mar;45(2):174-8.e1. [doi: 10.1016/j.jcjd.2020.07.005] [Medline: 33127288]
54. Eberle C, Loehnert M, Stichling S. Effectivness of specific mobile health applications (mHealth-apps) in gestational diabtetes mellitus: a systematic review. BMC Pregnancy Childbirth 2021 Dec 05;21(1):808 [FREE Full text] [doi: 10.1186/s12884-021-04274-7] [Medline: 34865645]
55. Leblalta B, Kebaili H, Lee S. PDB1 telemedicine use for gestational diabetes: a systematic review and meta-analysis of randomised controlled trials. Value in Health Regional Issues 2020 Sep;22:S32. [doi: 10.1016/j.vhri.2020.07.160]
56. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2015 Oct;110(1):1-9. [doi: 10.1016/j.diabres.2015.07.007] [Medline: 26264410]
57. Fantinelli S, Marchetti D, Verrocchio MC, Franzago M, Fulcheri M, Vitacolonna E. Assessment of psychological dimensions in telemedicine care for gestational diabetes mellitus: a systematic review of qualitative and quantitative studies. Front Psychol 2019;10:153 [FREE Full text] [doi: 10.3389/fpsyg.2019.00153] [Medline: 30804842]
58. 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]
59. Fleming PS, Koletsi D, Pandis N. Blinded by PRISMA: are systematic reviewers focusing on PRISMA and ignoring other guidelines? PLoS One 2014;9(5):e96407 [FREE Full text] [doi: 10.1371/journal.pone.0096407] [Medline: 24788774]
60. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
61. 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]
62. Garnweidner-Holme L, Hoel Andersen T, Sando MW, Noll J, Lukasse M. Health care professionals' attitudes toward, and experiences of using, a culture-sensitive smartphone app for women with gestational diabetes mellitus: qualitative study. JMIR Mhealth Uhealth 2018 May 14;6(5):e123 [FREE Full text] [doi: 10.2196/mhealth.9686] [Medline: 29759959]

63. Surendran S, Lim CS, Koh GC, Yew TW, Tai ES, Foong PS. Women's usage behavior and perceived usefulness with using a mobile health application for gestational diabetes mellitus: mixed-methods study. Int J Environ Res Public Health 2021 Jun 21;18(12):6670 [FREE Full text] [doi: 10.3390/ijerph18126670] [Medline: 34205744]
64. Bromuri S, Puricel S, Schumann R, Krampf J, Ruiz J, Schumacher M. An expert personal health system to monitor patients affected by gestational diabetes mellitus: a feasibility study. J Ambient Intelligence Smart Environ 2016 Mar 15;8(2):219-237. [doi: 10.3233/ais-160365]
65. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
66. Khalil C. Understanding the adoption and diffusion of a telemonitoring solution in gestational diabetes mellitus: qualitative study. JMIR Diabetes 2019 Nov 28;4(4):e13661 [FREE Full text] [doi: 10.2196/13661] [Medline: 31778118]
67. Johnson QB, Berry DC. Impacting diabetes self-management in women with gestational diabetes mellitus using short messaging reminders. J Am Assoc Nurse Pract 2018 Jun;30(6):320-326. [doi: 10.1097/JXX.0000000000000059] [Medline: 29878964]
68. Nicholson WK, Beckham AJ, Hatley K, Diamond M, Johnson L, Green SL, et al. The Gestational Diabetes Management System (GooDMomS): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. BMC Pregnancy Childbirth 2016 Sep 21;16(1):277 [FREE Full text] [doi: 10.1186/s12884-016-1064-z] [Medline: 27654119]
69. Peleg M, Shahar Y, Quaglini S, Broens T, Budasu R, Fung N, et al. Assessment of a personalized and distributed patient guidance system. Int J Med Inform 2017 May;101:108-130. [doi: 10.1016/j.ijmedinf.2017.02.010] [Medline: 28347441]
70. Peleg M, Shahar Y, Quaglini S, Fux A, García-Sáez G, Goldstein A, et al. MobiGuide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains. User Model User-Adap Inter 2017 Mar 11;27(2):159-213. [doi: 10.1007/s11257-017-9190-5]
71. Varnfield M, Redd C, Stoney RM, Higgins L, Scolari N, Warwick R, et al. M♡THer, an mHealth system to support women with gestational diabetes mellitus: feasibility and acceptability study. Diabetes Technol Ther 2021 May;23(5):358-366 [FREE Full text] [doi: 10.1089/dia.2020.0509] [Medline: 33210954]
72. Pais S, Parry D, Petrova K, Rowan J. Acceptance of using an ecosystem of mobile apps for use in diabetes clinic for self-management of gestational diabetes mellitus. Stud Health Technol Inform 2017;245:188-192. [Medline: 29295079]
73. Edwards KJ, Bradwell HL, Jones RB, Andrade J, Shawe JA. How do women with a history of gestational diabetes mellitus use mHealth during and after pregnancy? Qualitative exploration of women's views and experiences. Midwifery 2021 Jul;98:102995. [doi: 10.1016/j.midw.2021.102995] [Medline: 33784541]
74. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the feasibility of use of an online dietary assessment tool (myfood24) in women with gestational d.iabetes. Nutrients 2018 Aug 23;10(9):1147 [FREE Full text] [doi: 10.3390/nu10091147] [Medline: 30142898]
75. Rasekaba T, Nightingale H, Furler J, Lim WK, Triay J, Blackberry I. Women, clinician and IT staff perspectives on telehealth for enhanced gestational diabetes mellitus management in an Australian rural/regional setting. Rural Remote Health 2021 Jan;21(1):5983 [FREE Full text] [doi: 10.22605/RRH5983] [Medline: 33478229]
76. Pustozerov E, Popova P. Mobile-based decision support system for gestational diabetes mellitus. In: Proceedings of the 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). 2018 Presented at: 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT); May 07-08, 2018; Yekaterinburg, Russia. [doi: 10.1109/usbereit.2018.8384546]
77. Garnweidner-Holme LM, Borgen I, Garitano I, Noll J, Lukasse M. Designing and developing a mobile smartphone application for women with gestational diabetes mellitus followed-up at diabetes outpatient clinics in Norway. Healthcare (Basel) 2015 May 21;3(2):310-323 [FREE Full text] [doi: 10.3390/healthcare3020310] [Medline: 27417764]
78. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
79. Appendix E Evidence tables. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/cg143/evidence/appendix-e-evidence-tables-pdf-186634335 [accessed 2021-08-12]
80. CASP checklists. CASP. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2021-08-15]
81. Mixed Methods Appraisal Tool (MMAT) – Version 2011. McGill University. 2011. URL: https://tinyurl.com/mr3wjpd9 [accessed 2021-08-20]
82. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
83. Tigwell G, Menzies R, Flatla D. Designing for situational visual impairments: supporting early-career designers of mobile content. In: Proceedings of the 2018 Designing Interactive Systems Conference. 2018 Presented at: DIS '18: Designing Interactive Systems Conference 2018; Jun 9 - 13, 2018; Hong Kong China. [doi: 10.1145/3196709.3196760]

84. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. J Health Serv Res Policy 2005 Jul;10 Suppl 1:6-20. [doi: 10.1258/1355819054308576] [Medline: 16053580]
85. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
86. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
87. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
88. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
89. Yee LM, McGuire JM, Taylor SM, Niznik CM, Simon MA. "I was tired of all the sticking and poking": identifying barriers to diabetes self-care among low-income pregnant women. J Health Care Poor Underserved 2015 Aug;26(3):926-940. [doi: 10.1353/hpu.2015.0073] [Medline: 26320923]
90. Leziak K, Birch E, Jackson J, Strohbach A, Niznik C, Yee LM. Identifying mobile health technology experiences and preferences of low-income pregnant women with diabetes. J Diabetes Sci Technol 2021 Sep;15(5):1018-1026 [FREE Full text] [doi: 10.1177/1932296821993175] [Medline: 33605158]
91. Yee L, Taylor S, Young M, Williams M, Niznik C, Simon M. Evaluation of a text messaging intervention to support self-management of diabetes during pregnancy among low-income, minority women: qualitative study. JMIR Diabetes 2020 Aug 10;5(3):e17794 [FREE Full text] [doi: 10.2196/17794] [Medline: 32773367]
92. Anderson RM, Funnell MM, Fitzgerald JT, Marrero DG. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. Diabetes Care 2000 Jun;23(6):739-743. [doi: 10.2337/diacare.23.6.739] [Medline: 10840988]
93. Mcmillan B, Easton K, Delaney B, Mitchell C. Reducing the risk of progression from gestational diabetes to type 2 diabetes mellitus: women's perspectives on the content, research development and evaluation of a mobile health technology lifestyle intervention to support postnatal dietary and activity lifestyle change. Reducing the risk of type 2 diabetes mellitus in primary care after gestational diabetes. 2017 Jan. URL: https://tinyurl.com/23rw8yke [accessed 2022-10-21]
94. Sherman LE, Greenfield PM. Forging friendship, soliciting support: a mixed-method examination of message boards for pregnant teens and teen mothers. Comput Human Behav 2013 Jan;29(1):75-85. [doi: 10.1016/j.chb.2012.07.018]
95. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
96. Dasgupta K, Da Costa D, Pillay S, De Civita M, Gougeon R, Leong A, et al. Strategies to optimize participation in diabetes prevention programs following gestational diabetes: a focus group study. PLoS One 2013;8(7):e67878 [FREE Full text] [doi: 10.1371/journal.pone.0067878] [Medline: 23861824]
97. Vickery M, van Teijlingen E, Hundley V, Smith G, Way S, Westwood G. Midwives' views towards women using mHealth and eHealth to self-monitor their pregnancy: a systematic review of the literature. Eur J Midwifery 2020;4:36 [FREE Full text] [doi: 10.18332/ejm/126625] [Medline: 33537637]
98. Kruger DF, White K, Galpern A, Mann K, Massirio A, McLellan M, et al. Effect of modem transmission of blood glucose data on telephone consultation time, clinic work flow, and patient satisfaction for patients with gestational diabetes mellitus. J Am Acad Nurse Pract 2003 Aug;15(8):371-375. [doi: 10.1111/j.1745-7599.2003.tb00410.x] [Medline: 14509102]
99. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
100. Katz D, Price B, Holland S, Dalton N. Data, data everywhere,still too hard to link: ionsights from user interactions with diabetes apps. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3174077]
101. Fu HN, Rizvi RF, Wyman JF, Adam TJ. Usability evaluation of four top-rated commercially available diabetes apps for adults with type 2 diabetes. Comput Inform Nurs 2020 Jun;38(6):274-280 [FREE Full text] [doi: 10.1097/CIN.0000000000000596] [Medline: 31904594]
102. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]
103. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. J Biomed Inform 2016 Feb;59:115-129 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.008] [Medline: 26639894]

104. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
105. Jalil S, Myers T, Atkinson I. Design implications from the preliminary results of a telemedicine patient-technology interaction study. In: Proceedings of the 7th International Symposium on Visual Information Communication and Interaction. 2014 Presented at: VINCI '14: The 7th International Symposium on Visual Information Communication and Interaction; Aug 5 - 8, 2014; Sydney NSW Australia. [doi: 10.1145/2636240.2636860]
106. Freitas CM, Luzzardi PR, Cava RA, Winckler MA, Pimenta MS, Nedel LP. Evaluating usability of information visualization techniques. CNPq, FAPERGS, and CAPES. URL: https://dl.acm.org/doi/abs/10.1145/1556262.1556326 [accessed 2022-10-06]
107. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 09;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
108. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
109. Fontecha J, González I, Bravo J. A usability study of a mHealth system for diabetes self-management based on framework analysis and usability problem taxonomy methods. J Ambient Intell Human Comput 2019 Jun 24:1. [doi: 10.1007/s12652-019-01369-0]
110. Kontio E, Hyrkkänen U, Saarenpää T. Experiences on telemedicine solutions for diabetes care – Case eMedic project. In: Safe and Secure Cities. Cham: Springer; 2014.
111. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176 [FREE Full text] [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
112. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. A pilot study of a tool to stimulate physical activity in patients with COPD or type 2 diabetes in primary care. J Telemed Telecare 2014 Jan;20(1):29-34. [doi: 10.1177/1357633X13519057] [Medline: 24414397]
113. Nes AA, van Dulmen S, Eide E, Finset A, Kristjánsdóttir OB, Steen IS, et al. The development and feasibility of a web-based intervention with diaries and situational feedback via smartphone to support self-management in patients with diabetes type 2. Diabetes Res Clin Pract 2012 Sep;97(3):385-393. [doi: 10.1016/j.diabres.2012.04.019] [Medline: 22578890]
114. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
115. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
116. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
117. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integr Bioinform 2018 May 10;15(3):42-46 [FREE Full text] [doi: 10.1515/jib-2017-0030] [Medline: 29746254]
118. Wake DJ, He J, Czesak AM, Mughal F, Cunningham SG. MyDiabetesMyWay: an evolving national data driven diabetes self-management platform. J Diabetes Sci Technol 2016 Sep;10(5):1050-1058 [FREE Full text] [doi: 10.1177/1932296816648168] [Medline: 27162192]
119. Lapointe L, Rivard S. Getting physicians to accept new information technology: insights from case studies. CMAJ 2006 May 23;174(11):1573-1578 [FREE Full text] [doi: 10.1503/cmaj.050281] [Medline: 16717265]
120. Goodrich DE, Buis LR, Janney AW, Ditty MD, Krause CW, Zheng K, et al. Integrating an internet-mediated walking program into family medicine clinical practice: a pilot feasibility study. BMC Med Inform Decis Mak 2011 Jun 24;11:47 [FREE Full text] [doi: 10.1186/1472-6947-11-47] [Medline: 21702957]
121. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
122. Klaassen B, van Beijnum BJ, Hermens HJ. Usability in telemedicine systems-A literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
123. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]

Abbreviations

BG: blood glucose
GDM: gestational diabetes mellitus
HP: health professional
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SUS: system usability scale

Edited by R Kukafka; submitted 18.05.22; peer-reviewed by T Rasekaba, Y Shahar; comments to author 04.07.22; revised version received 16.07.22; accepted 28.09.22; published 27.10.22.

Please cite as:
Safiee L, Rough DJ, Whitford H
Barriers to and Facilitators of Using eHealth to Support Gestational Diabetes Mellitus Self-management: Systematic Literature Review of Perceptions of Health Care Professionals and Women With Gestational Diabetes Mellitus
J Med Internet Res 2022;24(10):e39689
URL: https://www.jmir.org/2022/10/e39689
doi:10.2196/39689
PMID:36301613

Review

The Current Status of Telemedicine Technology Use Across the World Health Organization European Region: An Overview of Systematic Reviews

Francesc Saigí-Rubió[1,2*], PhD; Israel Júnior Borges do Nascimento[3,4*], CPath; Noemí Robles[1], PhD; Keti Ivanovska[1], MA; Che Katz[1], MSc; Natasha Azzopardi-Muscat[5], PhD; David Novillo Ortiz[5], PhD

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

[2]Research Group in Epidemiology and Public Health in the Digital Health context, Universitat Oberta de Catalunya, Barcelona, Spain

[3]School of Medicine and University Hospital, Federal University of Minas Gerais, Belo Horizonte, Brazil

[4]School of Medicine, Medical College of Wisconsin, Milwaukee, WI, United States

[5]Division of Country Health Policies and Systems, World Health Organization, Regional Office for Europe, Copenhagen, Denmark

*these authors contributed equally

Corresponding Author:
David Novillo Ortiz, PhD
Division of Country Health Policies and Systems
World Health Organization, Regional Office for Europe
Marmorej 51
Copenhagen, 2100
Denmark
Phone: 45 61614868
Email: dnovillo@who.int

Abstract

Background: Several systematic reviews evaluating the use of telemedicine by clinicians, patients, and health authorities to improve the delivery of care in the 53 member states of the World Health Organization (WHO) European Region have been conducted in recent years. However, a study summarizing the findings of these reviews has not been conducted.

Objective: This overview of systematic reviews aimed to summarize findings regarding the use of telemedicine across the 53 member states and identify the medical fields and levels of care in and at which the effectiveness, feasibility, and applicability of telemedicine have been demonstrated. The barriers to and facilitators of telemedicine use were also evaluated and collated to help with the design and implementation of telemedicine interventions.

Methods: Through a comprehensive systematic evaluation of the published and unpublished literature, we extracted clinical, epidemiological, and technology-related data from each review included in the study. We focused on evaluating the barriers to and facilitators of the use of telemedicine apps across the 53 member states considered. We rated the methodological quality of each of the included reviews based on A Measurement Tool to Assess Systematic Review 2 approach and judged the overall certainty of evidence by using the Grading of Recommendations, Assessment, Development, and Evaluations methodology. The entire process was performed by 2 independent authors.

Results: This overview drew on data from >2239 primary studies, with >20,000 enrolled patients in total, within the WHO European Region. On the basis of data from randomized trials, observational studies, and economic evaluations from several countries, the results show a clear benefit of telemedicine technologies in the screening, diagnosis, management, treatment, and long-term follow-up of a series of chronic diseases. However, we were unable to pool the results into a reliable numeric parameter because of the high heterogeneity of intervention methodologies, scheduling, primary study design discrepancies, settings, and geographical locations. In addition to the clinical outcomes of the interventions, the social and economic outcomes are highlighted.

Conclusions: The application of telemedicine is well established across countries in the WHO European Region; however, some countries could still benefit from the many uses of these digital solutions. Barriers related to users, technology, and infrastructure were the largest. Conversely, the provision of health services using technological devices was found to significantly enhance patients' clinical outcomes, improve the long-term follow-up of patients by medical professionals, and offer logistical benefits for both patients and health workers.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42022309375; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=309375

(J Med Internet Res 2022;24(10):e40877) doi:10.2196/40877

KEYWORDS

telemedicine; Europe; World Health Organization; mobile phone

Introduction

Telemedicine is an accessible, cost-effective medical system, delivering high-quality care and reducing overall morbidity and mortality [1,2]. Telecommunications have benefited patient-related outcomes, improved health workers' performance, reduced health workers' workload, and decreased the isolation of health care professionals in remote locations [3,4]. Remote clinical care has increased, particularly during the COVID-19 pandemic [5-7]. The pandemic decreased in-person outpatient consultations and consequently increased telehealth legislation and public health guidance, which indirectly contributed to decline in transmissibility and mortality rates [8-10].

In the World Health Organization (WHO) European Region, an extensive body of literature has recently been produced, evidencing multiple positive health-related outcomes and the creation of integrated and finely structured remote health counseling programs [11-15]. European Union countries are covered by the European Commission's digital policies and priorities, which provide a common framework for digital interventions [16]. In addition, the European Commission provides funding programs to develop and implement these guidelines. No study has collated and summarized the available evidence to indicate the status of telemedicine in Europe. Therefore, this overview of systematic reviews aims to summarize findings regarding the use of telemedicine across the 53 member states of the WHO European Region and to identify the medical fields and levels of care in and at which the effectiveness, feasibility, and applicability of telemedicine have been demonstrated. The barriers to and facilitators of telemedicine use were also evaluated and collated to help with the design and implementation of telemedicine interventions.

Methods

Overview

The protocol for this overview of systematic reviews was published on February 17, 2022, in PROSPERO (International Prospective Register of Systematic Reviews; CRD42022309375; Multimedia Appendix 1). There were no substantial deviations from the proposed methodology. We adhered to an adapted version of the guidelines of the Cochrane Handbook for Systematic Reviews of Interventions and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and the Preferred Reporting Items for Overviews of Systematic Reviews checklist [17-19].

Ethical Considerations

This study relied on secondary data; therefore, no ethics approval or patient consent was required.

Search Strategy

Five databases (PubMed, Embase, Web of Science, the Cochrane Library, and Scopus) were searched from their inception to February 14, 2022. The search strategy is presented in Multimedia Appendix 2. All studies were processed in EndNote X9 (Clarivate) and subsequently imported into Covidence. In addition, we manually searched the first 2 pages of Google Scholar results and reviewed shortlisted records to identify additional studies. If a full-text study could not be obtained, a ResearchGate request was sent to gain full access.

Selection Criteria

Two investigators independently assessed titles and abstracts and analyzed appropriate studies through full-text evaluation. Cochrane and non-Cochrane systematic reviews with or without meta-analyses were included if they had adequately displayed the status of telemedicine among the 53 member states of the WHO European Region or reported on the barriers to and facilitators of the use of such technologies, regardless of publication data and the primary language. Reviews were considered eligible if >50% of the primary studies originated from the WHO European Region [20]. As telemedicine solutions and the publication time for manuscripts increased during the pandemic, we considered preprints and unpublished data. Exclusion criteria were (1) study unrelated to telemedicine, (2) study full text unavailable on the web, and (3) the scope of interventions does not include the WHO European Region. Any disagreement was resolved by discussion a third reviewer. We classified selected studies into systematic reviews (with or without meta-analyses), scoping reviews, and "others" (studies that were not classified into either of the 2 previously mentioned designs yet used a comprehensive execution methodology).

Data Extraction and Management

Two investigators independently extracted data by using Excel (Microsoft). A third party resolved discrepancies. The data extraction form (Multimedia Appendix 3) contains review identification features, telemedicine specialty, medical specialty or disease focus, countries and settings of focus, sample size, the main findings, barriers, facilitators, and the main challenges associated with the use of telemedicine.

Assessment of Methodological Quality

Two investigators independently appraised methodological quality by using A Measurement Tool to Assess Systematic Review 2 (AMSTAR 2). Discrepancies were resolved through consensus. In addition to the systematic reviews of intervention

trials, additional types of literature were included. Some AMSTAR 2 ratings were therefore adjusted (Multimedia Appendix 4 [21-53]). After rating each domain, overall confidence in the results of the review were judged as "critically low," "low," "moderate," or "high." Adherence ratings for the transparency of the researchers' judgments were reported, with explanations for each item.

Data Synthesis and Evaluation of the Level of Evidence

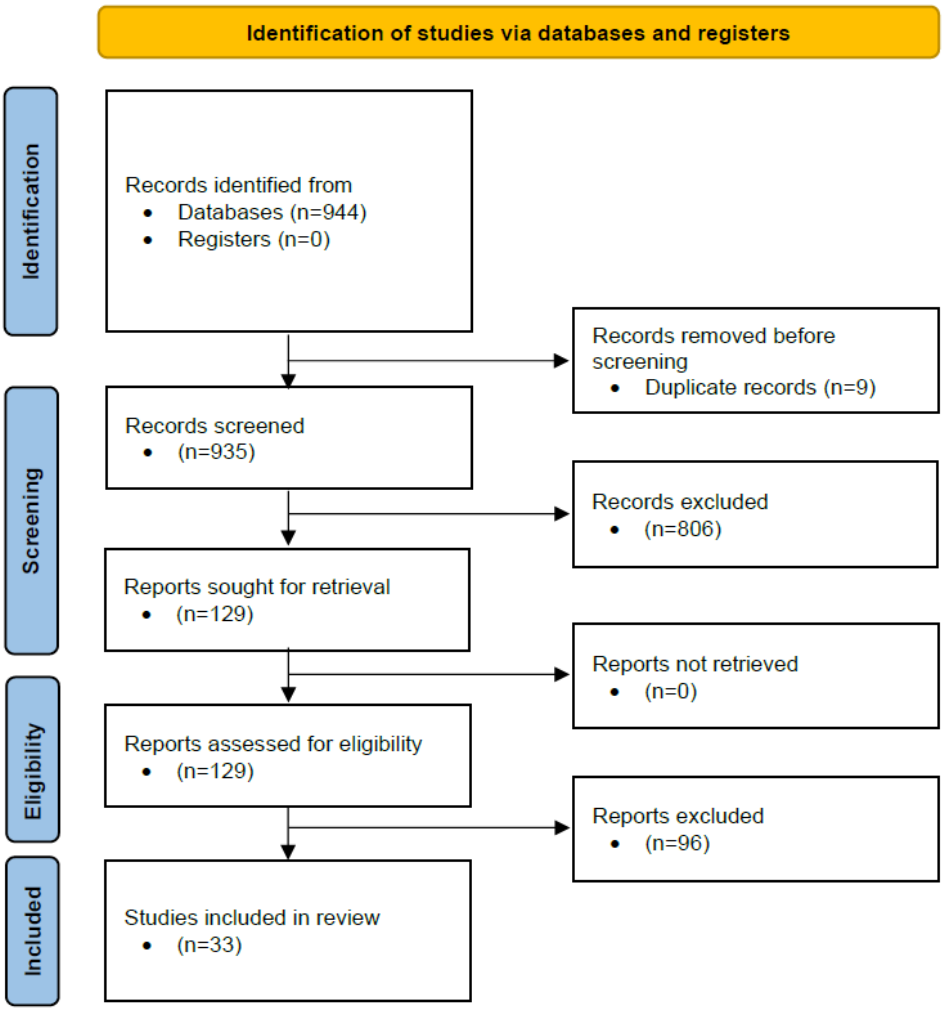
Evidence was synthesized based on the core disease or condition by using the *International Classification of Diseases*, 10th edition (ICD-10). A comprehensive narrative description of the characteristics and main findings was created and displayed in summarization tables. Furthermore, significant barriers and facilitators were presented, categorized, and discussed using the tree-mapping method, which displays hierarchical data as a set of nested rectangles. Limitations were also evaluated, and the effect of publication and small-study biases on results was considered. Finally, using an adapted version of the Grading of Recommendations, Assessment, Development, and Evaluations methodology, the evidence was assessed considering 5 modifiers: risk of bias in studies, inconsistency, imprecision, indirectness, and publication bias [54].

Results

Overview

In total, 944 records were retrieved, including 9 duplicates. In title and abstract screening, 806 publications were excluded. Of the remaining studies, 96 were excluded. Therefore, 33 articles were included in the final analysis. Additional records were found after checking the reference lists of included reviews. The overview flowchart is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for new systematic reviews that include searches of databases and registers only.



Characteristics of the Included Reviews

The characteristics of the included reviews are reported in in Multimedia Appendix 5 [21-53]. A total of 33 reviews were published between 2011 and 2022, mostly (17/33, 51%) between 2020 and 2021. The reviews were published in a range of indexed journals in English, Portuguese, or German. Translation support was sought through Cochrane TaskExchange. Regarding primary studies, 23 out of 53 member states of the European Region had at least one study evaluating the status of telemedicine (the United Kingdom, 17/33, 52%; Italy, 15/33, 45%; Denmark, 13/33, 39%; the Netherlands, 13/33, 39%;

Germany, 8/33, 24%; Norway, 8/33, 24%; Belgium, 6/33, 18%; Austria, Finland, France, and Sweden, 5/33, 15%; Spain, 4/33, 12%; Greece, 3/33, 9%; Iceland, Poland, Switzerland, and Türkiye, 2/33, 6%; and Albania, Ireland, Ukraine, Romania, and the Russian Federation, 1/33, 3%). The included reviews focused on various conditions or diseases, mainly those related to mental and behavioral disorders (Chapter V of ICD-10; 4/33, 12%), diseases of the circulatory system (Chapter IX of ICD-10; 4/33, 12%), diseases of the respiratory system (Chapter X of ICD-10; 4/33, 12%), diseases of the nervous system (Chapter VI of ICD-10; 3/33, 9%), and diseases of the skin and subcutaneous tissue (Chapter XII of ICD-10; 3/33, 9%). Of the 33 studies, 12 (33%) were classified as "multifocal studies" because they assessed multiple conditions or diseases and could not be assigned to just 1 chapter of the ICD-10.

Population and Study Designs

A total of 2239 primary studies were characterized as observational, interventional, medical, and economic modeling-based analyses and mixed methods studies. Publication designs were mostly "systematic reviews without meta-analyses" (19/33, 58%), "scoping reviews" (8/33, 24%), "others" (3/33, 9%), and "systematic reviews with meta-analyses" (3/33, 9%). Not all reviews specified the number of patients, but the data suggest that there were 61,589 patients.

Quality and Certainty of Evidence in Individual Systematic Reviews

The results of the AMSTAR 2 assessment showed that the main methodological weaknesses were a lack of protocol registration, no evaluation of the overall risk of bias by using validated approaches, a lack of disclosure and justification of excluded studies, and the absence of detailed reporting of the critical characteristics of the included reviews (Table 1). In summary, 88% (29/33) of the systematic reviews were judged to deliver "critically low" quality evidence and 12% (4/33) "low" quality evidence. None of the reviews produced high- or moderate-quality evidence. Therefore, confidence in the overall tendency of the effect was limited (Multimedia Appendix 6).

Table 1. Reliability of included reviews based on A Measurement Tool to Assess Systematic Review (AMSTAR 2) judgments[a].

Review ID (reference)	1[b]	2[c]	3[d]	4[e]	5[f]	6[g]	7[h]	8[i]	9[j]	10[k]	11[l]	12[m]	13[n]	14[o]	15[p]	16[q]	Overall quality
Allner et al [21]	Y[r]	PY[s]	N[t]	PY	Y	Y	N	N	N	N	NMAC[u]	NMAC	N	N	NMAC	Y	Very Low[v]
Brunetti al [22]	Y	N	Y	PY	Y	Y	N	N	Y	N	Y	N	N	N	Y	Y	Very Low[v]
Carbo et al [23]	Y	PY	Y	N	Y	Y	N	PY	Y	Y	Y	N	N	Y	N	Y	Very Low[v]
Cordes et al [24]	Y	N	Y	PY	N	N	N	PY	Y	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Cruz et al [25]	Y	N	Y	PY	Y	Y	N	Y	N	N	NMAC	NMAC	N	N	NMAC	Y	Low[w]
Elbaz et al [26]	Y	N	Y	PY	Y	N	N	PY	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Farabi et al [27]	Y	N	Y	PY	Y	Y	N	PY	Y	N	NMAC	NMAC	Y	N	NMAC	Y	Very Low[v]
Gaveikaite et al [28]	Y	N	Y	PY	Y	Y	N	Y	Y	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Glinkowski et al [29]	Y	N	Y	PY	Y	Y	N	N	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Hallensleben et al [30]	Y	N	Y	PY	Y	Y	N	Y	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Hartasanchez et al [31]	Y	N	Y	PY	Y	Y	N	PY	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Hrynyschyn et al [32]	Y	N	Y	PY	Y	Y	N	Y	Y	N	NMAC	NMAC	Y	N	NMAC	Y	Very Low[v]
Karamanidou et al [33]	Y	PY	Y	PY	Y	Y	N	PY	N	Y	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Kierkegaard et al [34]	Y	PY	Y	PY	Y	N	N	N	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Kingsdorf et al [35]	Y	PY	Y	PY	Y	Y	N	PY	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Labiris et al [36]	Y	N	Y	PY	N	N	N	PY	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Maresca et al [37]	Y	PY	Y	N	N	N	N	PY	N	N	NMAC	NMAC	N	N	NMAC	N	Very Low[v]
Martin et al [38]	Y	PY	Y	N	Y	Y	N	Y	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
McFarland et al [39]	Y	Y	Y	PY	Y	Y	N	Y	PY	N	Y	Y	Y	Y	Y	Y	Low[w]
Mold et al [40]	Y	PY	N	PY	Y	Y	Y	Y	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Nielsen et al [41]	Y	PY	Y	PY	N	N	N	PY	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
O'Cathail et al [42]	Y	PY	Y	PY	Y	Y	N	PY	N	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Ohannessian et al [43]	Y	N	Y	PY	Y	Y	N	PY	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Pron et al [44]	Y	PY	Y	PY	N	N	N	PY	PY	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]
Raja et al [45]	Y	N	Y	PY	Y	Y	N	PY	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Simmonds-Buckley et al [46]	Y	PY	Y	Y	Y	Y	N	Y	PY	N	Y	Y	Y	Y	Y	Y	Low[w]
Singh et al [47]	Y	N	Y	PY	N	N	N	Y	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Tokgoz et al [48]	Y	Y	Y	PY	Y	Y	N	PY	Y	Y	NMAC	NMAC	Y	Y	NMAC	Y	Low[w]
Trettel et al [49]	Y	N	N	PY	N	N	N	N	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Udsen et al [50]	Y	PY	Y	PY	Y	Y	N	Y	N	N	NMAC	NMAC	N	Y	NMAC	N	Very Low[v]
Verma et al [51]	Y	N	Y	PY	Y	Y	Y	N	N	Y	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Willard et al [52]	N	PY	N	PY	Y	Y	N	N	N	N	NMAC	NMAC	N	N	NMAC	Y	Very Low[v]
Zanin et al [53]	Y	PY	Y	PY	Y	Y	N	PY	PY	N	NMAC	NMAC	N	Y	NMAC	Y	Very Low[v]

[a]Judgments were made by 2 overview authors based on AMSTAR 2, a critical appraisal tool for systematic reviews that include randomized or nonrandomized studies of health care interventions or both.

[b]Domain 1—Did the research questions and inclusion criteria for the review include the components of PICO (Patients, Intervention, Comparator, and

Outcomes)

[c]Domain 2—Did the report of the review contain an explicit statement that the review methods were established before the conduct of the review and did the report justify any significant deviations from the protocol?

[d]Domain 3—Did the review authors explain their selection of the study designs for inclusion in the review?

[e]Domain 4—Did the review authors use a comprehensive literature search strategy?

[f]Domain 5—Did the review authors perform study selection in duplicate?

[g]Domain 6—Did the review authors perform data extraction in duplicate?

[h]Domain 7—Did the review authors provide a list of excluded studies and justify the exclusions?

[i]Domain 8—Did the review authors describe the included studies in adequate detail?

[j]Domain 9—Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?

[k]Domain 10—Did the review authors report on the sources of funding for the studies included in the review?

[l]Domain 11—If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?

[m]Domain 12—If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?

[n]Domain 13—Did the review authors account for RoB in individual studies when interpreting or discussing the results of the review?

[o]Domain 14—Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?

[p]Domain 15—If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small-study bias) and discuss its likely impact on the results of the review?

[q]Domain 16—Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

[r]Y: methodological requirements met.

[s]PY: methodological requirements partly met.

[t]N: methodological requirements not met.

[u]NMAC: no meta-analysis conducted.

[v]XX: studies rated as “critically low.”

[w]X: studies rated as “low.”

Systematic Review Findings

The identified interventions were mainly telephone- and videoconferencing-based methodologies, although they also included mobile apps and exchanges of medical test results (Multimedia Appendix 7 [21-53]). Most studies focused on the effectiveness of telemedicine interventions. The studies demonstrated that, as telemedicine was effective in reducing time to access treatment (1/33, 3%; Chapter IX of ICD-10), time for clinical decisions (1/33, 3%; Chapter IX of ICD-10), unnecessary repeated examinations (1/33, 3%; Chapter VII of ICD-10), length of stay in hospital (1/33, 3%; Chapter IX of ICD-10), number of emergency visits (1/33, 3%; Chapter X of ICD-10), and the number of false positives (1/33, 3%; Chapter VII of ICD-10) and also provided more accurate diagnoses (2/33, 6%; Chapter XII of ICD-10), it improved some clinical outcomes such as anxiety and depression in mental health disorders (2/33, 6%; Chapter V of ICD-10), neurological symptoms (1/33, 3%; Chapter VI of ICD-10), , exacerbation rates in patients with chronic obstructive pulmonary disease (1/33, 3%; Chapter X of ICD-10), wound healing time in some skin diseases (1/33, 3%; Chapter XII of ICD-10), and aphasia symptoms (1/33, 3%; Chapter XVIII of ICD-10). Moreover, telemedicine was reliable and sensitive for detecting changes in cognition over time (1/33, 3%; Chapter V of ICD-10) and improving patients’ quality of life (2/33, 6%; Chapter X of ICD-10; 1/33, 3%; Chapter VI of ICD-10; and 1/33, 3%; Chapter XXI of ICD-10) and quality-adjusted life years (1/33, 3%; Chapter IX of ICD-10).

Nonsignificant or inconclusive effects were found for other outcomes, such as mortality rates in circulatory and skin diseases (2/33, 6%; Chapter IX of ICD-10; 1/33, 3%; Chapter XII of ICD-10), the number of excisions in skin diseases (1/33, 3%; Chapter XII of ICD-10), and the number of hospital admissions (1/33, 3%; Chapter IX of ICD-10; 2/33, 6%; Chapter X of ICD-10).

A total of 5 studies evaluated the usability and acceptance of telemedicine by medical personnel and patients with multiple morbidities, as well as their satisfaction with it. High acceptability was primarily because of cost reduction compared with standard care, convenience, improved follow-up, adherence to planned treatment, and time-saving. A study reported telehealth’s cost-effectiveness, finding no statistically significant difference between standard care and telehealth care.

Barriers to and Facilitators of the Use of Telemedicine Interventions

The barriers, facilitators, and main challenges associated with the use of telemedicine are reported in Multimedia Appendix 8 [21-53]. Barriers and facilitators were grouped into the following domains: individual; organizational; clinical; economic; technological; and ethics, security, and privacy issues. Most barriers were in the individual domain, followed by technological; organizational; clinical; and ethics, security, and privacy issues domains, and finally the economic domain. Most facilitators were in the individual domain, followed by organizational and clinical domains, and then technological; economic; and ethics, security, and privacy issues domains (Table 2).

Table 2. List of barriers to and facilitators of the implementation of telemedicine across the 53 member states of the World Health Organization European Region and the main methodological limitations of the included studies.

Domain	Barriers	Facilitators
Individual domain	• Shortcomings in technology-related knowledge and skill [21,24,41,45,51] • Resistance to change [40,42,49,51] • Patients' age [24,40,41,45] • Lack of motivation or support [36,45,51] • Lack of confidence [49,51] • Challenges for individuals with disabilities [26,51] • Patients' preference for face-to-face consultations [40,41,51] • Low satisfaction [21,51] • Language barriers [41,51] • Lack of acceptance [52] • Lack of usefulness [41] • Less personal contact through telemedicine [39] • Invasiveness [39] • High attrition rate [35]	• Patient empowerment [31,35,51] • Participatory design [33,35,41,42] • Motivation and engagement [31,36,42,51] • Convenience [40,51,53] • Patients' age [24,41,45] • Trust in technology [35,45] • Patients feel safe and empowered to discuss personal issues [51] • Physicians' training and skills [31,45] • Satisfaction [36,51] • Adoption of digital culture [52] • Patients sharing their experiences [45]
Organizational domain	• The lack of integration into clinicians' workflows [31,42,44,51] • Socioeconomic aspects (financial limitations) [24,25,31,40] • Lack of access to a helpful caregiver [26,39,45] • Sociocultural aspects [21,40] • Increase in workload [51] • Scheduling conflicts [31] • Lack of governance [52] • No appropriate Health Information Systems framework [49] • Organizational issues creating barriers to long-term implementation [26]	• Reduction in response time [40,51,53] • Integration into clinicians' workflows [31,42] • Decrease in workload [37,51] • Access to a helpful caregiver and insights into patient's home environment [45,51] • Pandemic- created acceptance of technology [51] • Increased adherence [51] • Coordination between healthcare levels [52] • Telemedicine champions [34]
Clinical domain	• Limited scientific evidence [29,30,32,33,35] • Patient recruitment barriers and low rates of patient participation [28,46] • Difficulty in making clinical decisions [49,51] • Changes to consultation protocols [51] • Insufficient consultation time [51] • Loss of physical and visual assessment of symptoms [51]	• Clinical and professional benefits [33,34,51] • Assessment after a specified period with service evaluations, including feedback from key stakeholders [33,42] • Multidisciplinary care team interventions [33] • The establishment of guidelines [49] • Reduction in the number of visits [36] • Frequent and multimodal communication between the health care professional and patient [38] • Greater safety and efficacy [42] • Better monitoring of cases [51]
Economic domain	• Elevated cost of implementation [23,27,45] • Lack of funding model [34,42,43,51] • Scarce economic benefits [21]	• Financial framework [49,51] • Financial benefits [34,45] • Cost savings [36]
Technological domain	• Issues with internet access [24,26,31,35] • Technology needs further development [23,37,45] • Usability factors [25,31,41] • Issues with information technology and systems infrastructure [25,42] • Concerns about the reliability of the technology [21,41,45] • Issues surrounding infrastructure [44] • Conflicts of interoperability [52] • Difficulties in implementation and follow-up over a longer period [27] • Difficulties in readability [25] • Limited accessibility to electronic devices [24]	• Usability and user satisfaction factors [31,33,45] • Internet availability [31,35] • Possibilities of technology development [52] • Accessibility support [31,35] • Adaptable and self-configurable [52]
Ethics, security, and privacy issues	• Private data security concerns or issues [38,45,51] • Regulatory concerns or issues [21,44,49] • Concerns about patient and staff safety [21,41] • Ethical aspects [21]	• Legal framework [49]

Discussion

Principal Findings

This overview of systematic reviews shows a substantial and unprecedented collection of findings, as it included relevant data from >2239 primary studies, with >20,000 enrolled patients in total, within the WHO European Region. On the basis of data from observational studies, randomized trials, and economic evaluations from several European countries, the results showed a clear benefit of telemedicine interventions in the screening, diagnosis, management, treatment, and long-term follow-up of a range of clinically and epidemiologically significant diseases.

The telemedicine technological solutions addressed have proven to be valid, reliable, and accurate in providing faster access to expert advice, decreasing the number of unnecessary specialist referrals and in-office consultations, as well as increasing patient satisfaction experience. In a comprehensive literature review of studies from the United States, Canada, Brazil, and Australia, Liddy et al [55] reported an increasing number of medical specialties adopting innovative health solutions in daily practice. This overview of systematic reviews has highlighted the scientific priority of research in the evaluation of disease-related clinical, economic, and social outcomes, focusing on medical conditions considered chronic diseases, such as mental, cardiovascular, and respiratory diseases [56].

Most studies were concentrated in European countries (such as Germany, Italy, Spain, and the United Kingdom), while Eastern Europe (such as Albania, Croatia, and Ukraine) was not evaluated in any study. Countries developing digital health implementation must consider leadership, governance, strategy, investment, infrastructure, legislation, policy, compliance, workforce, services, and apps in their digital health strategies. The 2015 WHO Global Survey on eHealth [57] revealed that 38% of the member states had not developed national telehealth policies or strategies, and 49% did not have mHealth programs [58,59].

Several studies reported barriers and facilitators that should be considered when planning and implementing telemedicine interventions. The individual domain was found to be the most influential in the use of telemedicine interventions, giving place to a greater number of barriers and facilitators. Shortcoming in technology-related knowledge and skills was the main challenge cited, followed by health care professionals' resistance to procedural change [60-62]. According to some studies, the lack of technological applications integrated into clinicians' practice exacerbated this, resulting in scheduling conflicts and affecting the quality of delivery. The presence of clinician champions working alongside other health care professionals might promote service adoption [63,64].

In this overview, health care professionals had heavy workloads that seemed to influence resistance by overshadowing benefits [61,65]. Conversely, integrating telemedicine into clinicians' workflows, establishing guidelines, and increasing coordination among the levels of care were organizational changes required for the proper adoption of telemedicine [66]. In addition, training and skills mitigated the shortcomings in knowledge and skills, enabling health care professionals to use telemedicine easily [67,68]. These potential barriers should be identified early in the process of planning the implementation of changes. Physicians' and patients' needs, characteristics, acceptance, and satisfaction must be further assessed through research informed by the technology acceptance model [69], unified theory of acceptance and use of technology [70], theory of planned behavior [71], and theory of organization and environment [72] and so too must the reliability, usefulness, and ease of use of technologies [73]. This will enable the formulation of strategies to avoid resistance to change [74].

Many clinical factors have been shown to influence the success of telemedicine in the WHO European Region, mainly the lack of definitive scientific evidence on its clinical contribution. Others included management, care delivery, and outcomes for a particular pathology. More research was considered necessary to provide evidence of both the clinical benefits of telemedicine and improved case monitoring [75].

Telemedicine resistance was reported as often being due to patients' lack of confidence, lack of motivation or support, or sociocultural aspects [76-78]. A feeling of having less personal contact with the clinician, lack of access to a helpful caregiver, and face-to-face preference hindered patients' use of telemedicine [79-81], but shortcomings in technology-related knowledge and skills posed the main challenge [62,82]. Conversely, patient motivation, engagement, and empowerment were considered the main facilitators, enabling patients to use telemedicine easily [78,83-85]. Access to a helpful caregiver, insight into the home environment, and adoption of digital culture might reduce resistance, as the pandemic has shown [86]. Patients who trusted technology and were satisfied with web-based consultations showed no resistance to telemedicine, felt safe and empowered to discuss personal issues, and had experiences similar to those with face-to-face consultations. However, better response time was one of the largest facilitators of telemedicine. Web-based consultation also promoted increased adherence, indicating a correlation between telemedicine compliance and convenience [87,88].

However, patients with disabilities or older patients encountered difficulties when using telemedicine [89], which increased their reluctance to use it. Patients' age was also considered a potential facilitator, especially among younger individuals [90]. Frequent and multimodal communication between health care professionals and patients, as well as patients sharing their experiences, might reduce the aforementioned difficulties [91,92].

Access to funding and the high costs associated with implementation were economic barriers. Similarly, socioeconomic aspects emerged as obstacles to the functional integration of telemedicine apps. The implementation of a financial framework must be considered. However, outcomes were positive when technology was financially beneficial [93-95].

Internet access, technology development, usability, infrastructure, and interoperability were the main barriers to telemedicine intervention delivery, usability, and user

satisfaction, while the availability of technology development was a mediator and facilitator [96].

The most common barriers associated with ethics, security, and privacy issues were privacy and data security and data-related regulatory concerns [97,98]. Health care professionals and patients also raised concerns about safety, especially with mobile medical apps. According to the 2015 WHO Global Survey on eHealth, 80% of the member states had laws to protect individual health data, but 53% had none in place to allow individuals to access their own data; only 43% had policies or legislation regarding medical jurisdiction, liability, or compensation [58,79]. However, only 1 study highlighted the need to establish a legal framework to ensure that new telemedicine technologies complied with the constitution, legislation, regulations, and existing contracts [49]. The WHO European Region and European Commission have focused their policies on data exchange and regulatory aspects and now offer a series of frameworks and recommendations that national health plans should include to ensure the success of telemedicine and digital health as a whole. These are the cases of the European Health Information Initiative fostered by the WHO European Region, which aims to harmonize the health information gathered in European countries, and the European Commission's European data strategy, which promotes the creation of a single market for data, including health data [99]. However, other relevant aspects, such as clinical, organizational, and human factors, have either been disregarded or do not have a clear direction. This scenario poses a considerable challenge for the formulation of public policies and strategies by health care institutions, where decisions on telemedicine use should not be overlooked.

Finally, based on the solid effectiveness telemedicine technologies can deliver, policy makers and stakeholders should not only facilitate the implementation of these applications but also recognize and tackle drawbacks to maximize the likelihood of use success. Research is confronted with the challenge of producing such evidence, a prerequisite for the generalized adoption of telemedicine. Nevertheless, none of the included studies reached "moderate" or "high" reporting quality based on the AMSTAR 2 methodology. Studies have rarely reported items considered critical for assessing the methodological quality of systematic reviews. The existence of such reporting inappropriateness significantly affected our results, as the overall quality of the evidence was directly affected by the overall limited reporting quality of the included reviews. Notably, several other evidence makers have emphasized the occurrence of systematic reviews with poor or very poor reporting completeness [100,101]. As a partial solution to this issue, we strongly suggest the need to adhere to the basic principles used among high-quality evidence researchers, with considerable attention paid to critical features (protocol registration before project initiation, appropriate search strategy and literature search, rationale for excluding studies, risk of bias appraisal of included studies, appropriateness of meta-analytical methods [when pertinent], consideration of risk of bias in interpreting review findings, and assessment of publication bias). Thus, by appropriately using core reporting features, systematic reviews (and consequently, overviews of systematic reviews) can guide decisions on accurate, succinct, credible, and comprehensive summaries of the best available evidence on a topic.

Limitations

A total of 5 databases were explored, focusing only on systematic reviews, meta-analyses, and bibliometric analyses, thus limiting the exhaustivity of the search. Furthermore, although we initially identified almost 1000 studies for screening, our overview found only 33 reviews meeting our inclusion and exclusion criteria. Consequently, the representativeness of our findings can be questioned considering the number of primarily identified records. However, despite using a highly sensitive search strategy, designed with collaboration between a field specialist and librarian, the "over retrieval" of records might not only associate with wrong selection of identifiers and keywords by systematic review authors but also reflect indexation issues. In addition, this could also reflect the absence of a reliable description of methods used throughout study execution (resulting in the exclusion of shortlisted records) and the scarcity of investigations on this particular subject of study. The information sources were peer-reviewed publications; therefore, some relevant information from other sources (eg, gray literature) may have been missed. Lower quality scores based on AMSTAR 2 may have reflected incomplete reports rather than unqualified review methods, such as some aspects not considered by the authors; for example, a lack of protocol registration or clarity on the characteristics of the included and excluded studies.

Conclusions

The results underscore the need to design dynamic approaches for telemedicine interventions in the WHO European Region. Potential barriers should be identified early in the process. The barriers and facilitators identified in this overview, as well as their influence, should be further investigated because only clear evidence will support the formulation of strategies to avoid resistance to change [74]. Poorer nations should also be included to benefit from emerging health technologies and to avoid geoeconomic research bias [102-104]. The WHO European Region and European Commission have developed several initiatives to foster the development and implementation of telemedicine. These include some that are more general, such as the inclusion of telemedicine and digital health as a key aspect in their policy frameworks (eg, Global Strategy on Digital Health 2020-2025 by the WHO) and others that are more focused on implementation (Horizon 2020 and Horizon Europe funding programs and the European Reference Networks) [99]. The WHO European Region will continue leveraging the potential of telemedicine in the context of the Digital Health Action Plan for the WHO European Region (2023-2030), which was adopted in September 2022. In the context of these policy frameworks, these initiatives recognize not only the power of telemedicine to break down geographical barriers and expand access to health services but also the need for mechanisms to mitigate barriers and risks.

Disclaimer

DNO and NA-M are staff members of the World Health Organization. The authors alone are responsible for the views expressed in this paper, and they do not necessarily represent the decisions, policies, or views of the World Health Organization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol for this overview of systematic reviews published in PROSPERO (International Prospective Register of Systematic Reviews).
[PDF File (Adobe PDF File), 160 KB - jmir_v24i10e40877_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 15 KB - jmir_v24i10e40877_app2.docx]

Multimedia Appendix 3
Data extraction form.
[DOCX File , 16 KB - jmir_v24i10e40877_app3.docx]

Multimedia Appendix 4
A Measurement Tool to Assess Systematic Review 2 judgment for each included study.
[DOCX File , 773 KB - jmir_v24i10e40877_app4.docx]

Multimedia Appendix 5
Main characteristics of included reviews.
[DOCX File , 25 KB - jmir_v24i10e40877_app5.docx]

Multimedia Appendix 6
Summary of findings table for main outcomes.
[DOCX File , 19 KB - jmir_v24i10e40877_app6.docx]

Multimedia Appendix 7
Main findings from included reviews.
[DOCX File , 43 KB - jmir_v24i10e40877_app7.docx]

Multimedia Appendix 8
List of barriers, facilitators, limitations, and current challenges.
[DOCX File , 41 KB - jmir_v24i10e40877_app8.docx]

References

1. Armaignac DL, Saxena A, Rubens M, Valle CA, Williams LM, Veledar E, et al. Impact of telemedicine on mortality, length of stay, and cost among patients in progressive care units: experience from a large healthcare system. Crit Care Med 2018 May;46(5):728-735 [FREE Full text] [doi: 10.1097/CCM.0000000000002994] [Medline: 29384782]
2. 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]
3. Telemedicine: Opportunities and developments in Member State. World Health Organization. Geneve, Switzerland: World Health Organization; 2010. URL: https://www.afro.who.int/publications/telemedicine-opportunities-and-developments-member-state [accessed 2022-10-17]
4. The Lancet Gastroenterology Hepatology. The potential of telemedicine in digestive diseases. Lancet Gastroenterol Hepatol 2019 Mar;4(3):185. [doi: 10.1016/S2468-1253(18)30359-5] [Medline: 30739659]
5. 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]

6. Suran M. Increased use of medicare telehealth during the pandemic. JAMA 2022 Jan 25;327(4):313. [doi: 10.1001/jama.2021.23332] [Medline: 35076683]
7. 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]
8. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
9. Casariego-Vales E, Blanco-López R, Rosón-Calvo B, Suárez-Gil R, Santos-Guerra F, Dobao-Feijoo M, On Behalf Of The Telea-Covid Lugo Comanagement Team. Efficacy of telemedicine and telemonitoring in at-home monitoring of patients with COVID-19. J Clin Med 2021 Jun 29;10(13):2893 [FREE Full text] [doi: 10.3390/jcm10132893] [Medline: 34209725]
10. Chauhan V, Galwankar S, Arquilla B, Garg M, Somma SD, El-Menyar A, et al. Novel coronavirus (COVID-19): leveraging telemedicine to optimize care while minimizing exposures and viral transmission. J Emerg Trauma Shock 2020;13(1):20-24 [FREE Full text] [doi: 10.4103/JETS.JETS_32_20] [Medline: 32308272]
11. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
12. Purnomo J, Coote K, Mao L, Fan L, Gold J, Ahmad R, et al. Using eHealth to engage and retain priority populations in the HIV treatment and care cascade in the Asia-Pacific region: a systematic review of literature. BMC Infect Dis 2018 Feb 17;18(1):82 [FREE Full text] [doi: 10.1186/s12879-018-2972-5] [Medline: 29454322]
13. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
14. Escobar-Viera CG, Cernuzzi LC, Miller RS, Rodríguez-Marín HJ, Vieta E, González Toñánez M, et al. Feasibility of mHealth interventions for depressive symptoms in Latin America: a systematic review. Int Rev Psychiatry 2021 May;33(3):300-311 [FREE Full text] [doi: 10.1080/09540261.2021.1887822] [Medline: 34102945]
15. Jiménez-Molina Á, Franco P, Martínez V, Martínez P, Rojas G, Araya R. Internet-based interventions for the prevention and treatment of mental disorders in Latin America: a scoping review. Front Psychiatry 2019 Sep 13;10:664 [FREE Full text] [doi: 10.3389/fpsyt.2019.00664] [Medline: 31572242]
16. A Europe fit for the digital age: Empowering people with a new generation of technologies. European Commission. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en [accessed 2022-05-09]
17. Bougioukas KI, Liakos A, Tsapas A, Ntzani E, Haidich AB. Preferred reporting items for overviews of systematic reviews including harms checklist: a pilot tool to be used for balanced reporting of benefits and harms. J Clin Epidemiol 2018 Jan;93:9-24. [doi: 10.1016/j.jclinepi.2017.10.002] [Medline: 29037888]
18. 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]
19. Pollock M, Fernandes R, Becker L, Pieper D, Hartling L. Chapter V: Overviews of reviews. In: Higgins JP, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: Cochrane Collaboration; 2011.
20. Countries. World Health Organization. URL: https://www.who.int/countries [accessed 2022-08-10]
21. Allner R, Wilfling D, Kidholm K, Steinhäuser J. Telemedicine projects in rural areas of Germany. A systematic evaluation with the "Model for Assessment of Telemedicine". Z Evid Fortbild Qual Gesundhwes 2019 May;141-142:89-95. [doi: 10.1016/j.zefq.2019.03.005] [Medline: 30962133]
22. Brunetti ND, De Gennaro L, Correale M, Santoro F, Caldarola P, Gaglione A, et al. Pre-hospital electrocardiogram triage with telemedicine near halves time to treatment in STEMI: a meta-analysis and meta-regression analysis of non-randomized studies. Int J Cardiol 2017 Apr 01;232:5-11. [doi: 10.1016/j.ijcard.2017.01.055] [Medline: 28089154]
23. Carbo A, Gupta M, Tamariz L, Palacio A, Levis S, Nemeth Z, et al. Mobile technologies for managing heart failure: a systematic review and meta-analysis. Telemed J E Health 2018 Dec 1;24(12):958-968. [doi: 10.1089/tmj.2017.0269] [Medline: 29608430]
24. Cordes L, Loukanova S, Forstner J. Scoping review of the effectiveness of screen-to-screen-therapy compared to face-to-face-therapy on naming performance for patients with aphasia. Z Evid Fortbild Qual Gesundhwes 2020 Nov;156-157:1-8 [FREE Full text] [doi: 10.1016/j.zefq.2020.08.002] [Medline: 33032962]
25. Cruz J, Brooks D, Marques A. Home telemonitoring in COPD: a systematic review of methodologies and patients' adherence. Int J Med Inform 2014 Apr;83(4):249-263. [doi: 10.1016/j.ijmedinf.2014.01.008] [Medline: 24529402]
26. Elbaz S, Cinalioglu K, Sekhon K, Gruber J, Rigas C, Bodenstein K, et al. A systematic review of telemedicine for older adults with dementia during COVID-19: an alternative to in-person health services? Front Neurol 2021 Dec 14;12:761965 [FREE Full text] [doi: 10.3389/fneur.2021.761965] [Medline: 34970210]

27. Farabi H, Rezapour A, Jahangiri R, Jafari A, Rashki Kemmak A, Nikjoo S. Economic evaluation of the utilization of telemedicine for patients with cardiovascular disease: a systematic review. Heart Fail Rev 2020 Nov;25(6):1063-1075. [doi: 10.1007/s10741-019-09864-4] [Medline: 31760593]
28. Gaveikaite V, Grundstrom C, Winter S, Chouvarda I, Maglaveras N, Priori R. A systematic map and in-depth review of European telehealth interventions efficacy for chronic obstructive pulmonary disease. Respir Med 2019;158:78-88 [FREE Full text] [doi: 10.1016/j.rmed.2019.09.005] [Medline: 31614305]
29. Glinkowski WM, Karlińska M, Karliński M, Krupiński EA. Telemedicine and eHealth in Poland from 1995 to 2015. Adv Clin Exp Med 2018 Feb;27(2):277-282 [FREE Full text] [doi: 10.17219/acem/74124] [Medline: 29521073]
30. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. Int J Chron Obstruct Pulmon Dis 2019;14:1681-1690 [FREE Full text] [doi: 10.2147/COPD.S207187] [Medline: 31440044]
31. Hartasanchez SA, Heen AF, Kunneman M, García-Bautista A, Hargraves IG, Prokop LJ, et al. Remote shared decision making through telemedicine: a systematic review of the literature. Patient Educ Couns 2022 Feb;105(2):356-365. [doi: 10.1016/j.pec.2021.06.012] [Medline: 34147314]
32. Hrynyschyn R, Dockweiler C, Iltner J, Hornberg C. Teleconsultation for vascular- and diabetes-associated chronic wounds: a systematic review of health-related and economic implications. Hautarzt 2020 Feb;71(2):114-123. [doi: 10.1007/s00105-019-04498-x] [Medline: 31659390]
33. Karamanidou C, Natsiavas P, Koumakis L, Marias K, Schera F, Schäfer M, et al. Electronic patient-reported outcome-based interventions for palliative cancer care: a systematic and mapping review. JCO Clin Cancer Inform 2020 Jul;4:647-656 [FREE Full text] [doi: 10.1200/CCI.20.00015] [Medline: 32697604]
34. Kierkegaard P. Mapping telemedicine efforts: surveying regional initiatives in Denmark. Telemed J E Health 2015 May;21(5):427-435. [doi: 10.1089/tmj.2014.0123] [Medline: 25664377]
35. Kingsdorf S, Pančocha K. Looking at Europe's recent behavioral telehealth practices for children and families impacted by neurodevelopmental disabilities. Int J Dev Disabil (forthcoming) 2021 May 17:1-16. [doi: 10.1080/20473869.2021.1925403]
36. Labiris G, Panagiotopoulou EK, Kozobolis VP. A systematic review of teleophthalmological studies in Europe. Int J Ophthalmol 2018 Feb 18;11(2):314-325 [FREE Full text] [doi: 10.18240/ijo.2018.02.22] [Medline: 29487825]
37. Maresca G, Maggio MG, De Luca R, Manuli A, Tonin P, Pignolo L, et al. Tele-neuro-rehabilitation in Italy: state of the art and future perspectives. Front Neurol 2020 Sep 30;11:563375 [FREE Full text] [doi: 10.3389/fneur.2020.563375] [Medline: 33101176]
38. Martin S, Sutcliffe P, Griffiths F, Sturt J, Powell J, Adams A, et al. Effectiveness and impact of networked communication interventions in young people with mental health conditions: a systematic review. Patient Educ Couns 2011 Nov;85(2):e108-e119. [doi: 10.1016/j.pec.2010.11.014] [Medline: 21239133]
39. 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]
40. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
41. Nielsen AS, Kidholm K, Kayser L. Patients' reasons for non-use of digital patient-reported outcome concepts: a scoping review. Health Informatics J 2020 Dec;26(4):2811-2833 [FREE Full text] [doi: 10.1177/1460458220942649] [Medline: 32731773]
42. O'Cathail M, Sivanandan MA, Diver C, Patel P, Christian J. The use of patient-facing teleconsultations in the national health service: scoping review. JMIR Med Inform 2020 Mar 16;8(3):e15380 [FREE Full text] [doi: 10.2196/15380] [Medline: 32175911]
43. Ohannessian R, Schott AM, Colin C, Nighoghossian N, Medeiros de Bustos E, Moulin T. Acute telestroke in France: a systematic review. Rev Neurol (Paris) 2020 May;176(5):316-324. [doi: 10.1016/j.neurol.2019.11.004] [Medline: 32147201]
44. Pron G, Ieraci L, Kaulback K, Medical Advisory Secretariat, Health Quality Ontario. Internet-based device-assisted remote monitoring of cardiovascular implantable electronic devices: an evidence-based analysis. Ont Health Technol Assess Ser 2012;12(1):1-86 [FREE Full text] [Medline: 23074419]
45. Raja M, Bjerkan J, Kymre IG, Galvin KT, Uhrenfeldt L. Telehealth and digital developments in society that persons 75 years and older in European countries have been part of: a scoping review. BMC Health Serv Res 2021 Oct 26;21(1):1157 [FREE Full text] [doi: 10.1186/s12913-021-07154-0] [Medline: 34696789]
46. Simmonds-Buckley M, Bennion MR, Kellett S, Millings A, Hardy GE, Moore RK. Acceptability and effectiveness of NHS-recommended e-therapies for depression, anxiety, and stress: meta-analysis. J Med Internet Res 2020 Oct 28;22(10):e17049 [FREE Full text] [doi: 10.2196/17049] [Medline: 33112238]
47. Singh TP, Vangaveti VN, Kennedy RL, Malabu UH. Role of telehealth in diabetic foot ulcer management - a systematic review. Aust J Rural Health 2016 Aug;24(4):224-229. [doi: 10.1111/ajr.12284] [Medline: 27098271]
48. Tokgöz P, Dockweiler C. Telemedicine in rehabilitation aftercare for women with breast cancer - a systematic literature review. Rehabilitation (Stuttg) 2022 Feb;61(1):17-24. [doi: 10.1055/a-1480-3712] [Medline: 34399436]

49. Trettel A, Eissing L, Augustin M. Telemedicine in dermatology: findings and experiences worldwide - a systematic literature review. J Eur Acad Dermatol Venereol 2018 Feb;32(2):215-224. [doi: 10.1111/jdv.14341] [Medline: 28516492]
50. Udsen FW, Hejlesen O, Ehlers LH. A systematic review of the cost and cost-effectiveness of telehealth for patients suffering from chronic obstructive pulmonary disease. J Telemed Telecare 2014 Jun;20(4):212-220. [doi: 10.1177/1357633X14533896] [Medline: 24803277]
51. Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. BJGP Open 2022 Jun;6(2):BJGPO.2021.0192 [FREE Full text] [doi: 10.3399/BJGPO.2021.0192] [Medline: 35031558]
52. Willard S, van Rossum E, Spreeuwenberg M, de Witte L. A typology of online care platforms for community-dwelling older adults in the Netherlands: a scoping review. Gerontechnology 2019 Sep 28;18(3):122-141. [doi: 10.4017/gt.2019.18.3.001.00]
53. Zanin E, Aiello EN, Diana L, Fusi G, Bonato M, Niang A, Italian working group on tele-neuropsychology (TELA). Tele-neuropsychological assessment tools in Italy: a systematic review on psychometric properties and usability. Neurol Sci 2022 Jan;43(1):125-138 [FREE Full text] [doi: 10.1007/s10072-021-05719-9] [Medline: 34751849]
54. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
55. Liddy C, Moroz I, Mihan A, Nawar N, Keely E. A systematic review of asynchronous, provider-to-provider, electronic consultation services to improve access to specialty care available worldwide. Telemed J E Health 2019 Mar;25(3):184-198. [doi: 10.1089/tmj.2018.0005] [Medline: 29927711]
56. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
57. 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]
58. From innovation to implementation: eHealth in the WHO European Region. Regional Office for Europe, World Health Organization. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/handle/10665/326317 [accessed 2022-05-11]
59. 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):941-50.e1 [FREE Full text] [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
60. Rabanifar N, Abdi K. Barriers and challenges of implementing telerehabilitation: a systematic review. Iran Rehabil J 2021 Jun 01;19(2):121-128. [doi: 10.32598/irj.19.2.1404.1]
61. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016 Aug 1;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
62. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
63. Loscertales FR, Rubió FS. Facilitators in the implantation of telemedicine services. Perspective of professionals involved in its design and implementation. An Sist Sanit Navar 2011;34(2):235-244 [FREE Full text] [doi: 10.4321/s1137-66272011000200009] [Medline: 21904405]
64. Payán DD, Frehn JL, Garcia L, Tierney AA, Rodriguez HP. Telemedicine implementation and use in community health centers during COVID-19: clinic personnel and patient perspectives. SSM Qual Res Health 2022 Dec;2:100054 [FREE Full text] [doi: 10.1016/j.ssmqr.2022.100054] [Medline: 35169769]
65. Østervang C, Vestergaard LV, Dieperink KB, Danbjørg DB. Patient rounds with video-consulted relatives: qualitative study on possibilities and barriers from the perspective of healthcare providers. J Med Internet Res 2019 Mar 25;21(3):e12584 [FREE Full text] [doi: 10.2196/12584] [Medline: 30907746]
66. Vindrola-Padros C, Singh KE, Sidhu MS, Georghiou T, Sherlaw-Johnson C, Tomini SM, et al. Remote home monitoring (virtual wards) for confirmed or suspected COVID-19 patients: a rapid systematic review. EClinicalMedicine 2021 Jul;37:100965 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100965] [Medline: 34179736]
67. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021 Dec 09;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]
68. Datta N, Derenne J, Sanders M, Lock JD. Telehealth transition in a comprehensive care unit for eating disorders: challenges and long-term benefits. Int J Eat Disord 2020 Nov;53(11):1774-1779. [doi: 10.1002/eat.23348] [Medline: 32715512]
69. 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]
70. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]

71. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
72. Kuan KK, Chau PY. A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. Inf Manag 2001 Oct;38(8):507-521. [doi: 10.1016/s0378-7206(01)00073-8]
73. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
74. Ross J, Stevenson F, Lau R, Murray E. Exploring the challenges of implementing e-health: a protocol for an update of a systematic review of reviews. BMJ Open 2015 Apr 13;5(4):e006773 [FREE Full text] [doi: 10.1136/bmjopen-2014-006773] [Medline: 25869685]
75. Borycki E. Quality and safety in eHealth: the need to build the evidence base. J Med Internet Res 2019 Dec 19;21(12):e16689 [FREE Full text] [doi: 10.2196/16689] [Medline: 31855183]
76. Stelson EA, Carr BG, Golden KE, Martin N, Richmond TS, Delgado MK, et al. Perceptions of family participation in intensive care unit rounds and telemedicine: a qualitative assessment. Am J Crit Care 2016 Sep;25(5):440-447. [doi: 10.4037/ajcc2016465] [Medline: 27587425]
77. Rising KL, Ricco JC, Printz AD, Hoon Woo S, Hollander JE. Virtual rounds: observational study of a new service connecting family members remotely to inpatient rounds. Gen Int Med Clin Innov 2016;1(3):44-47. [doi: 10.15761/GIMCI.1000113]
78. Mason AN, Brown M, Mason K. Telemedicine patient satisfaction dimensions moderated by patient demographics. Healthcare (Basel) 2022 Jun 01;10(6):1029 [FREE Full text] [doi: 10.3390/healthcare10061029] [Medline: 35742080]
79. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
80. Eslami Jahromi M, Ahmadian L. Evaluating satisfaction of patients with stutter regarding the tele-speech therapy method and infrastructure. Int J Med Inform 2018 Jul;115:128-133. [doi: 10.1016/j.ijmedinf.2018.03.004] [Medline: 29779715]
81. Lee PS, Koo S, Panter S. The value of physical examination in the era of telemedicine. J R Coll Physicians Edinb 2021 Mar;51(1):85-90. [doi: 10.4997/JRCPE.2021.122] [Medline: 33877145]
82. Robben SH, Perry M, Huisjes M, van Nieuwenhuijzen L, Schers HJ, van Weel C, et al. Implementation of an innovative web-based conference table for community-dwelling frail older people, their informal caregivers and professionals: a process evaluation. BMC Health Serv Res 2012 Aug 15;12:251 [FREE Full text] [doi: 10.1186/1472-6963-12-251] [Medline: 22894654]
83. Kruse CS, Lee K, Watson JB, Lobo LG, Stoppelmoor AG, Oyibo SE. Measures of effectiveness, efficiency, and quality of telemedicine in the management of alcohol abuse, addiction, and rehabilitation: systematic review. J Med Internet Res 2020 Jan 31;22(1):e13252 [FREE Full text] [doi: 10.2196/13252] [Medline: 32012048]
84. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
85. Ehlers DK, Huberty JL, de Vreede GJ. Can an evidence-based book club intervention delivered via a tablet computer improve physical activity in middle-aged women? Telemed J E Health 2015 Feb;21(2):125-131. [doi: 10.1089/tmj.2013.0360] [Medline: 25526014]
86. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic-a systematic review. Int J Environ Res Public Health 2022 May 17;19(10):6113 [FREE Full text] [doi: 10.3390/ijerph19106113] [Medline: 35627650]
87. Mason AN. The most important telemedicine patient satisfaction dimension: patient-centered care. Telemed J E Health 2022 Aug;28(8):1206-1214. [doi: 10.1089/tmj.2021.0322] [Medline: 34882032]
88. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]
89. Annaswamy TM, Verduzco-Gutierrez M, Frieden L. Telemedicine barriers and challenges for persons with disabilities: COVID-19 and beyond. Disabil Health J 2020 Oct;13(4):100973 [FREE Full text] [doi: 10.1016/j.dhjo.2020.100973] [Medline: 32703737]
90. Solans O, Vidal-Alaball J, Roig Cabo P, Mora N, Coma E, Bonet Simó JM, et al. Characteristics of citizens and their use of teleconsultations in primary care in the Catalan public health system before and during the COVID-19 pandemic: retrospective descriptive cross-sectional study. J Med Internet Res 2021 May 27;23(5):e28629 [FREE Full text] [doi: 10.2196/28629] [Medline: 33970867]
91. Krysta K, Romańczyk M, Diefenbacher A, Krzystanek M. Telemedicine treatment and care for patients with intellectual disability. Int J Environ Res Public Health 2021 Feb 11;18(4):1746 [FREE Full text] [doi: 10.3390/ijerph18041746] [Medline: 33670152]
92. Farsi D, Martinez-Menchaca HR, Ahmed M, Farsi N. Social media and health care (Part II): narrative review of social media use by patients. J Med Internet Res 2022 Jan 07;24(1):e30379 [FREE Full text] [doi: 10.2196/30379] [Medline: 34994706]

93. Senel E, Bas Y. Evolution of telepathology: a comprehensive analysis of global telepathology literature between 1986 and 2017. Turk Patoloji Derg 2020;36(3):218-226 [FREE Full text] [doi: 10.5146/tjpath.2019.01484] [Medline: 32282053]
94. Burg F, Pscherer A, Opitz OG. Digital communication in visceral medicine: regulatory framework for digital communication. Visc Med 2021 Oct 20;96(6):1-7 [FREE Full text] [doi: 10.1159/000519359] [Medline: 34934767]
95. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
96. Cernadas Ramos A, Barral Buceta B, Fernández da Silva Á, Bouzas-Lorenzo R, Garaikoetxea Iturriria A. The present and future of eHealth in Spain from a health management perspective. Int J Health Serv 2022 Jul;52(3):400-409. [doi: 10.1177/0020731420914836] [Medline: 32237961]
97. Jalali MS, Landman A, Gordon WJ. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
98. Solimini R, Busardò FP, Gibelli F, Sirignano A, Ricci G. Ethical and legal challenges of telemedicine in the era of the COVID-19 pandemic. Medicina (Kaunas) 2021 Nov 30;57(12):1314 [FREE Full text] [doi: 10.3390/medicina57121314] [Medline: 34946259]
99. Cioti AM, Stanescu AM, Grajdeanu IV, Serban B, Popescu E, Bratu OG, et al. Telemedicine in Europe - current status and future perspectives. Rev Med Mod 2019 Dec 30;26(4):165-168 [FREE Full text] [doi: 10.31689/rmm.2019.26.4.165]
100. Pagkalidou E, Anastasilakis DA, Kokkali S, Doundoulakis I, Tsapas A, Dardavessis T, et al. Reporting completeness in abstracts of systematic reviews of diagnostic test accuracy studies in cardiovascular diseases is suboptimal. Hellenic J Cardiol 2022 Feb 15;65:25-34 [FREE Full text] [doi: 10.1016/j.hjc.2022.02.001] [Medline: 35181563]
101. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. BMJ 2014 Jun 25;348:g3804 [FREE Full text] [doi: 10.1136/bmj.g3804] [Medline: 24965222]
102. Acharya KP, Pathak S. Applied research in low-income countries: why and how? Front Res Metr Anal 2019 Nov 14;4:3 [FREE Full text] [doi: 10.3389/frma.2019.00003]
103. Rochmyaningsih D. The developing world needs basic research too. Nature 2016 Jun 02;534(7605):7. [doi: 10.1038/534007a] [Medline: 27251238]
104. Amarante V, Burger R, Chelwa G, Cockburn J, Kassouf A, McKay A, et al. Underrepresentation of developing country researchers in development research. Appl Econ Lett 2022;29(17):1659-1664. [doi: 10.1080/13504851.2021.1965528]

Abbreviations

AMSTAR 2: A Measurement Tool to Assess Systematic Review 2
ICD-10: International Classification of Diseases, 10th edition
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
PROSPERO: International Prospective Register of Systematic Reviews
WHO: World Health Organization

Edited by G Eysenbach, T Leung; submitted 08.07.22; peer-reviewed by J Hron, S Xu; comments to author 02.08.22; revised version received 16.08.22; accepted 31.08.22; published 27.10.22.

Please cite as:
Saigí-Rubió F, Borges do Nascimento IJ, Robles N, Ivanovska K, Katz C, Azzopardi-Muscat N, Novillo Ortiz D
The Current Status of Telemedicine Technology Use Across the World Health Organization European Region: An Overview of Systematic Reviews
J Med Internet Res 2022;24(10):e40877
URL: https://www.jmir.org/2022/10/e40877
doi:10.2196/40877
PMID:36301602

Review

Examining Mental Workload Relating to Digital Health Technologies in Health Care: Systematic Review

Lisanne Kremer[1], MSc; Myriam Lipprandt[2], PhD; Rainer Röhrig[2], Prof Dr; Bernhard Breil[1], Prof Dr

[1]Faculty of Health Care, Niederrhein University of Applied Sciences, Krefeld, Germany

[2]University Hospital Rheinisch-Westfälische Technische Hochschule Aachen, Institute of Medical Informatics, Aachen, Germany

Corresponding Author:
Lisanne Kremer, MSc
Faculty of Health Care
Niederrhein University of Applied Sciences
Reinarzstr. 49
Krefeld, 47805
Germany
Phone: 49 2151 8226678
Email: Lisanne.Kremer@hs-niederrhein.de

Abstract

Background: The workload in health care is increasing and hence, mental health issues are on the rise among health care professionals (HCPs). The digitization of patient care could be related to the increase in stress levels. It remains unclear whether the health information system or systems and digital health technologies (DHTs) being used in health care relieve the professionals or whether they represent a further burden. The mental construct that best describes this burden of technologies is mental workload (MWL). The measurement methods of MWL are particularly relevant in this sensitive setting.

Objective: This review aimed to address 2 different but related objectives: identifying the factors that contribute to the MWL of HCPs when using DHT and examining and exploring the applied assessments for the measurement of MWL with a special focus on eye tracking.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement, we conducted a systematic review and processed a literature search in the following databases: MEDLINE (PubMed), Web of Science, Academic Search Premier and CINAHL (EBSCO), and PsycINFO. Studies were eligible if they assessed the MWL of HCPs related to DHT. The review was conducted as per the following steps: literature search, article selection, data extraction, quality assessment (using the Standard Quality Assessment Criteria for Evaluation Primary Research Papers From a Variety of Fields [QualSyst]), data analysis, and data synthesis (narrative and tabular). The process was performed by 2 reviewers (in cases of disagreement, a third reviewer was involved).

Results: The literature search process resulted in 25 studies that fit the inclusion criteria and examined the MWL of health care workers resulting from the use of DHT in health care settings. Most studies had sample sizes of 10-50 participants, were conducted in the laboratory, and had quasi-experimental or cross-sectional designs. The main results can be grouped into two categories: assessment methods and factors related to DHT that contribute to MWL. Most studies applied subjective methods for the assessment of MWL. Eye tracking did not play a major role in the selected studies. The factors contributing to a higher MWL were clustered into organizational and systemic factors.

Conclusions: Our review of 25 papers shows a diverse assessment approach toward the MWL of HCPs related to DHT as well as 2 groups of relevant contributing factors to MWL. Our results are limited in terms of interpretability and causality due to methodological weaknesses of the included studies and may be limited by some shortcomings in the search process. Future research should concentrate on adequate assessments of the MWL of HCPs dependent on the setting, the evaluation of quality criteria, and further assessment of the contributing factors to MWL.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42021233271; https://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42021233271

(J Med Internet Res 2022;24(10):e40946) doi:10.2196/40946

KEYWORDS

mental workload; mental workload measurement; assessment; health care professional; health information system(s); digital health technology; systematic review

Introduction

Background

The decrease in nursing staff with the simultaneous increase in patients with multiple morbidities in need of care means an increase in workload of the remaining nursing staff. The digitization of health care in theory should help to counteract this change and its consequences. However, in Germany in particular, the process is proceeding very slowly; Germany is ranked 16th out of 17 countries in the Bertelsmann Digital Health Index [1]. The application of digital health technology (DHT) is an important factor in the digitalization process. DHTs in the context of this review means technologies that are directly linked to outpatient and inpatient care and are implemented by nurses or physicians. By DHT, we mean, for instance, health information systems (HISs), medical devices, and other digital applications that support patient care from the perspective of health care professionals (HCPs).

In addition to the positive effects of the use of DHT, there is also evidence which suggests that its use can cause extra workload [2] and can consequently have a negative impact on HCPs' health [3]. However, it remains ambiguous which factors are specifically responsible for a high mental workload (MWL) during the use of DHT. Initial results show that this may be because of a lack of usability and user involvement as well as poor implementation processes [4,5].

Poor usability and other factors rooted in technology can cause a high MWL [5]. High workloads can cause errors independent of the operators status (novice or expert). Those errors often results form decision-making processes. [6]. When working with patients, however, susceptibility to errors as well as indecisiveness cannot be an option. Working in outpatient and inpatient care can be considered as working in safety-critical environments. Many tasks, varying in complexity, occur within limited time windows. Decisions could be supported by different DHTs through the structured and standardized presentation of information.

The interaction between the users and the systems is complex and interdependent, which contributes to difficulties in the prediction of effects related to the systems on the users [7].

Wickens et al [8] give a good practical example for this effect. During surgery, different complex tasks have to be performed by the surgeon in addition to observing the patient. In the event of a sudden change in the patient's vital signs, which can be potentially life-threatening, the surgeon has to promptly take an appropriate decision on how to proceed. Complex demands could result in an overload if they exceed the capacity of attentional resources [7]. Consequences of overload are an increasing vulnerability to errors and decreasing performance. In addition to serious consequences for patients, an overload also has drastic effects on employees. High workloads caused by several factors (including technology) result in consequences regarding the workers' health; technostress, mental health issues such as depression or burnout, and decreased job satisfaction are only a few of the alarming effects [9]. There is growing evidence that DHTs are contributing to increasing mental health problems, (eg, burnout of health care workers [10,11]). The investigation of MWL in different situations is a possible approach toward identifying the main causes behind, for example, emerging incidences of burnout in physicians and nurses [12].

Mental Workload

MWL can be defined using different approaches and is usually influenced by different and multiple factors. It is multidimensional and multifaceted and is one of the most important variables for understanding and predicting human performance.

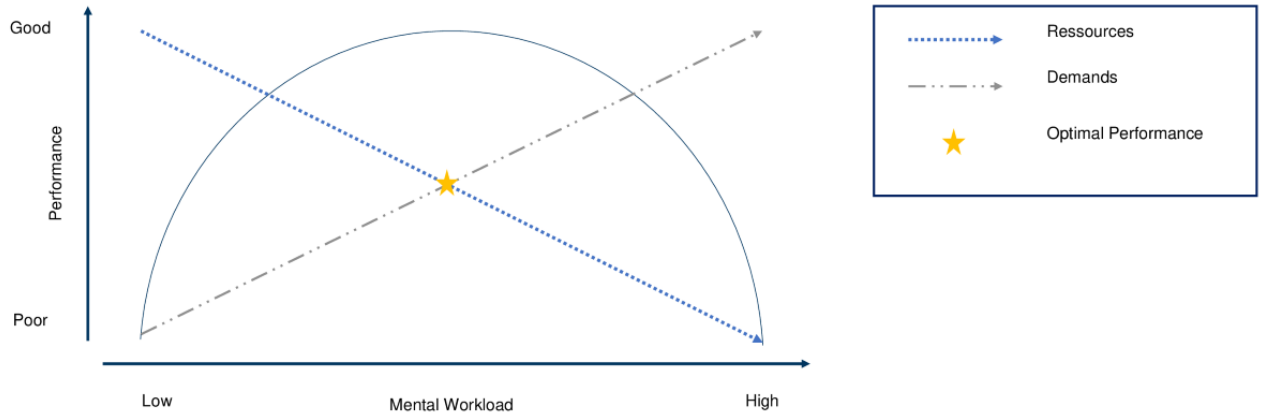
The possible definitional approaches of workload can be derived from two different perspectives:

1. MWL as an external variable referring to task requirements: the amount of work and the number of tasks to be completed (in a limited time), that is,
   task load
2. Interaction between task and human resources resulting in a subjective psychological experience [13,14]

Eggemeier et al [15] define MWL as the "proportion of the operator's information processing capacity or resources that is actually required to meet system demands." Gopher and Donchin [16] state that "mental workload may be viewed as the difference between capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time." They define MWL as a latent variable relating to the interaction between the operator and the task. As per Proctor [6], the definition of MWL is "a task [that] represents the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and past experience."

In summary, there is no all-encompassing, universally accepted definition of MWL. We define MWL as a construct that addresses the influence of task demands on operator resources resulting in an impact on psychological factors such as performance (Figure 1) but not in the sense of stress or acceptance.

Figure 1. Task demands and limited resources result in different workloads and performance aspects. The optimal performance can be reached when resources and demands are balanced and the level of mental workload is moderate. Figure 1 is based on the representation of the Yerkes-Dodsen law [17].



Especially during work, inadequate workload results in poorer performance [18]. Following the above definitions, a high workload can either be caused by unsuitable task requirements or by limited resources that are available in a cognitive manner, for example certain parts of the brain. The aim of measuring MWL is to determine the tasks and work processes that cause adverse or inappropriate levels of demands to draw conclusions about user performance as well as error prevention. Furthermore, the measurement of MWL can help identify factors that cause consequences such as technostress or burnout among nurses and physicians [10].

Assessment of Mental Workload

MWL assessment was first developed and applied in other safety-critical environments such as aviation or aerospace or nuclear power plants. Owing to similar conditions—already described—in the sociotechnical system, workload assessment is also a useful approach in the clinical setting.

The assessment of MWL can be performed by different techniques. A distinction between analytical and empirical methods may be drawn. Analytical methods tend to be used in system development, while empirical methods are used when workload is to be measured directly in the executing system or in the simulation [13].

Analytical assessment methods are simulation models, expert opinions, or task analyses. Empirical methods are distinguished into three different categories: performance measures, subjective methods, and physiological techniques [6]. Performance measures refer to the measures of the primary and a secondary task.

Depending on the situation and the underlying question, one or more of these techniques are appropriate to apply. Several factors should be considered when selecting assessments, including sensitivity, diagnostic ability, intrusiveness, validity, reliability, simplicity of use, and user acceptance [19].

Tao et al [20] analyzed the physiological assessment of MWL across different application areas. One main result was that MWL assessments were not essentially valid in all areas, for example, for all tasks and differed in their validity.

Charles and Nixon [21] provide an overview of physiological measures that discriminate between different MWL levels. They detect varying ranges in the sensitivity of these measures but provide an evidence base for their deployment.

These reviews concentrate on physiological measures, not on all possible assessments. Although physiological measures are gaining relevance in the field of MWL assessment, methods that can be applied quickly and easily can still probably be helpful, especially in the health care sector.

Objectives

The workload in health care institutions is high. A possible factor contributing to high workloads could be the use of DHT.

MWL is possibly the construct that can reflect best the workload caused by technologies. There is only light evidence for causes of MWL related to DHT. One reason might be that the health care sector has not been in the spotlight for researchers of human factors until now. To our knowledge, there currently is no review of the measurement methods for MWL caused by DHT.

As a primary objective, this systematic review intends to identify the impact of digital technologies, particularly HIS, on the workload of health care workers.

There are specific reviews investigating physiological methods assessing MWL as well as several papers studying the MWL in health care in general. We aimed to present a broader approach by looking at all methods that were used in the defined field while providing a more specific approach in focusing on DHT in particular, thus differing from already existing reviews to this topic [20,21]. We concentrated on a review of applied methods as well as their quality criteria. In addition (as secondary objectives), we aimed to assess what methods are being applied in health care to measure MWL relating to DHT. In particular, the application of eye tracking or pupillometry as a measurement method was investigated.

The research questions for this study are as follows:

1. In what manner do DHT contribute to the overall MWL of health care workers and which aspects or factors of DHT contribute to an increase in MWL?

2. What are the methods or assessments being applied to measure MWL related to HIS or digital technologies?
   - What role does eye tracking or pupillometry play in context of measurement?
   - What outcomes are being assessed via eye tracking?

Rationale

Many different factors have led to a significant increase in workload in the health care sector in the past few years [22]. Work-related stress has become one of the main challenges in the health care sector [23]. Nurses in particular report high levels of work-related stress that lead to negative physical and psychological effects for them as well as for their patients [24]. Many nurses describe themselves as feeling empty and report depressive symptoms [25,26]. In Germany in particular, the number of days of sick leave taken by nurses is increasing every year. In addition to musculoskeletal diseases, which account for the majority of sick leaves, absences because of mental illness are increasing significantly [27]. The past two years (2020-2021) brought about many other challenges as well.

Methods

Study Registration

This systematic review is registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42021233271) and follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [28].

Eligibility Criteria

We defined the inclusion criteria for this systematic review according to the population, intervention, comparison, outcome, context scheme and the corresponding research question or questions. The inclusion criteria related to the study population, measurement type (intervention or comparison), the outcome of the study, and the study setting (context). An additional inclusion criterion related to study design.

Study Design

This systematic review comprises 2 research questions. For both of these, we have included randomized controlled trials (RCTs), quasi-RCTs, case-control studies, and comparative cross-sectional studies as well as longitudinal design studies that either compare measurement methods for question 1 or generally measure MWL in the context of HISs and DHT.

Study Participants

We focused on HCPs who worked with DHTs that are directly related to patient care. These can be nurses, physicians, radiology assistants, medical students, or other clinicians. It is essential that the participants are supported by the HIS or DHT in their daily work with patients. We excluded studies that focused only on patients' views on DHT use.

Intervention or Measurement

We included studies measuring MWL related to DHT that were directly related to patient care. The studies should have investigated whether there is a direct or indirect effect of DHT on workers' MWL. Because the second research question evaluates the extent to which eye tracking is commonly used as a measurement method, we put a special focus on the inclusion of studies that apply eye tracking.

Study Setting

All types of study designs reporting original primary data as well as systematic reviews that adhered to our other inclusion criteria were included. We excluded commentaries, letters, and guidelines as well as scoping and narrative reviews.

Exclusion Criteria

We excluded studies that focused on the measurement of MWL in other contexts than health care (eg, aviation) as well as studies that were related to the measurement of allied constructs such as technostress or that focused on sources of MWL in health care other than DHT. In addition, we did not include studies that examined the workload of patients.

Outcomes

The primary outcome of this systematic review was to analyze the influence of DHT on the MWL of HCPs and medical or nursing students.

Secondary outcomes included the types of assessments that are applied to measure MWL related to DHT. Additionally, we examined the impact of eye tracking on the measurement of MWL related to DHT.

Information Sources

The following databases were systematically searched between January 20, 2021, and February 28, 2021, by using defined keywords (and synonyms) such as "mental workload," "health information system," "assessment," "health care professionals" and "eye tracking" that result in specified search strings (the block chain is shown in Multimedia Appendix 1): MEDLINE (PubMed), Web of Science, Academic Search Premier and CINAHL (EBSCO), and PsycINFO. In addition, we searched for relevant research in the reference sections of included studies as well as of relevant recently published reviews. The keywords were defined by reviewing thesaurus systems such as Medical Subject Headings, expert opinions, and reviews of relevant studies.

We updated our search in February 2022 by replicating this process.

Following PRISMA, we organized the search terms by database and research question in a separate document [28]. We have attached this document (Multimedia Appendix 2)

Search Strategy

The search strategy included the following four categories, each represented by keywords and synonyms: technologies used (eg, HIS), population (eg, HCPs), methods (eg, assessment), and MWL. In addition, eye tracking was added for research questions 2.1. and 2.2. The terms are linked by the Boolean operators AND or.

We restricted our search to articles published in the period between 2000 and 2022. This search time frame was chosen because it documents the development of the current generation

of prehospital communication technology, such as telemedicine and electronic patient care reports [29]. The literature search was limited to articles written in English or German, as both reviewers were sufficiently proficient in these languages.

Study Records

Data Management

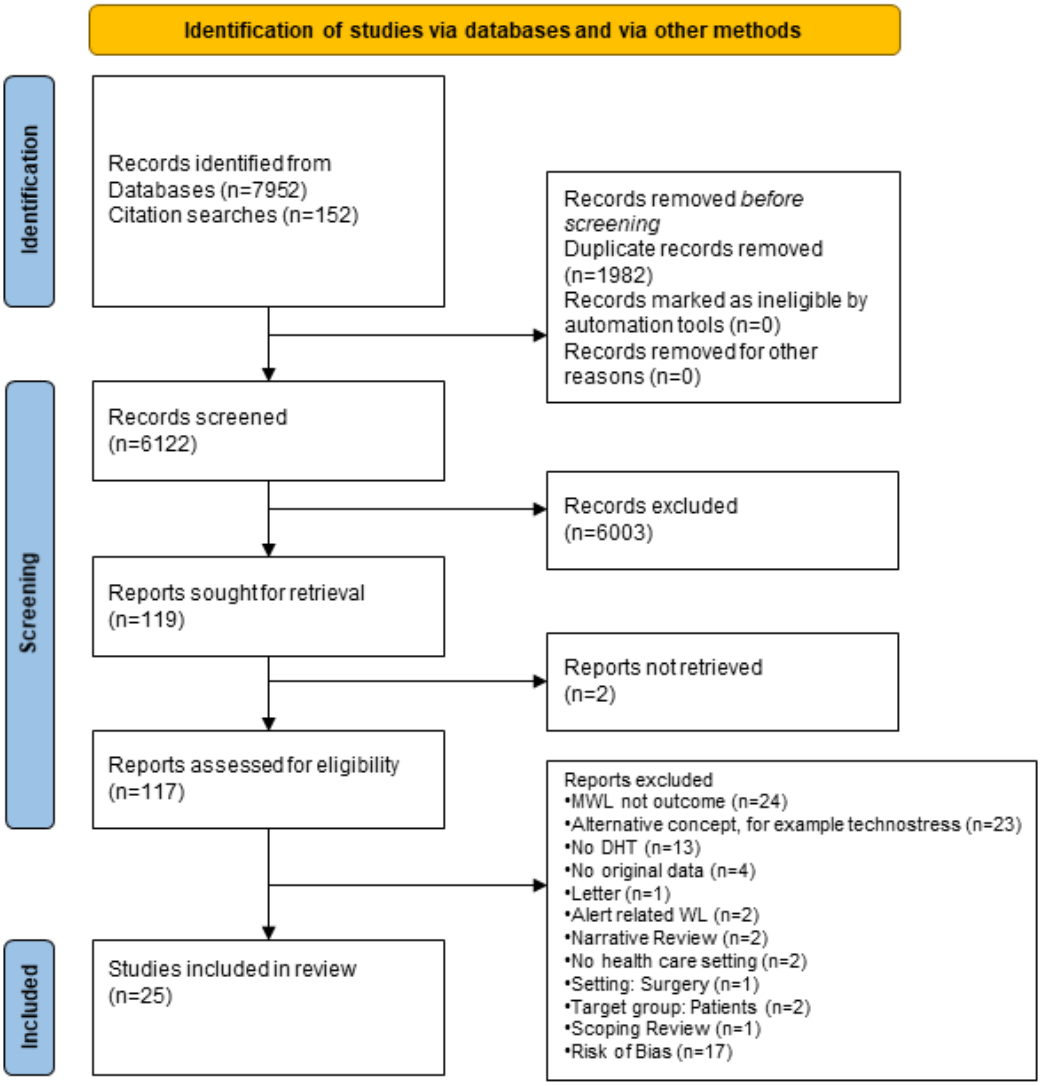
Citavi (Citavi 6 for Windows–Campus; QRS International) was used for literature handling, that is, importing of articles and further screening of the literature. The *Rayaan* web-based screening tool was used to support further abstract screening and full-text analysis in a structured format [30]. In this context, the inclusion and exclusion criteria were also provided, functioning as the basis for the analysis process. The included articles were then imported to an extraction sheet.

Selection Process

The selection process was performed by two reviewers, LK and BB, (and two conciliating reviewers, ML and RR) according to PRISMA guidelines and is displayed using a flowchart (Figure 2) First, both reviewers assessed the studies regarding the inclusion and exclusion criteria for abstract screening. We included studies that (1) focused on DHTs such as HISs that are directly related to patient care, (2) focused on the MWL of HCPs that is related to DHT/HIS, (3) assessed MWL or cognitive load related to DHT, and (4) were processed in a health care context. We excluded studies that (1) focused on the assessment of MWL in other contexts (eg, aviation), (2) were related to the assessment of allied constructs such as technostress (3) that focused on patients (relating to either technologies or workload) (4) that focused on MWL not related to DHT and (5) were nonoriginal works (letters, guidelines, and narrative reviews) and books. In the next step, the full texts of the resulting studies were assessed independently.

Finally, we searched the references of the papers for further possibly eligible studies. In case of disagreements in any of the phases, a discussion between the two reviewers (LK and BB) based on the inclusion criteria was first attempted. If the discussion turned out to be inconclusive, a third reviewer (ML and RR) was involved.

Figure 2. The figure displays the Flow Chart of the Search strategy starting with 8104 articles and resulting in 25 included studies. Most studies were excluded because MWL was not the primary outcome or the study focused alternative concepts [29]. DHT: digital health technology; MWL: mental workload.



Data Collection Process

A tabular extraction sheet for data extraction was used based on the outcomes of the review. To ensure uniformity across reviewers, we conducted a pretest standardization exercise before starting the data extraction process. Each reviewer extracted the themes of interest to an extraction sheet.

Risk of Bias in Individual Studies

Two evaluators independently rated the quality of the identified studies using the QualSyst Scale [31]. Disagreements were resolved via discussion (among LK and BB) or, if necessary, resolved by a third reviewer (ML and RR).

Studies were rated using a structured tool (comprising 14 items). If a study completely fulfilled a criterion, it was assigned 2 points. In case of partial fulfillment, 1 point was assigned. If the criterion was not fulfilled by the study, no point was assigned to the study.

If a criterion was not applicable to the study presented (eg, blinding of the investigator), it was removed from the assessment. The achieved points as a percentage of the possible total points were evaluated as per the following criteria: a score of <0.5 by both reviewers resulted in exclusion, studies with scores between 0.5 and 0.65 were classified as having a moderate risk of bias, and studies with scores >0.65 were classified as having a low risk of bias.

Data Items

LL and BB read the full texts and extracted information concerning identified and relevant aspects of the studies. We differentiated main study characteristics, measurements, and outcomes from relevant findings and recommendations.

In addition to the descriptive presentation of study characteristics and findings, we aimed to extract factors or aspects of DHT that contributed to an increase in MWL. Furthermore, we extracted information on how the included studies assessed the workload and in which settings eye tracking was used with regard to specific outcomes. On the basis of this, we developed an overview of the methods that can be used to measure MWL caused by DHTs meaningfully and validly. Furthermore, we assessed the studies concerning the categories of types of DHT and factors that contribute to a lower MWL.

The methods, settings, and outcomes were organized into logical categories that were rated by the reviewers. The typical categories of methods referring to MWL assessments were analytical or empirical techniques. Typical categories for settings were laboratory or field. Categories referring to assessed outcomes have to be defined during the reviewing process. In each category, we extracted how often an indicator for a category was applied (eg, category % = method applied/N studies) and how often combinations of specific indicators were used (eg, for total percentage with method A with setting B and outcome C, total % = combination applied/N studies). A typical indicator for category methods would be a questionnaire or subjective method. If an indicator was identified, the reviewers filled in the row with a 1; if no indicator was identified, for example, if the method was not applied, the table was filled in with a 0.

Data Analysis and Synthesis

After the initial screening of the search results, we did not conduct a meta-analysis because the results and quantifications of the measures varied widely. Instead, we performed descriptive analysis to summarize the data, in which we first compared the studies in terms of the evaluation methods used (qualitative, quantitative, or mixed methods) and then performed a comparison of their survey methods.

We used the following two nonquantitative approaches for data synthesis: tabulation and a narrative approach.

In a first step, all main characteristics of each study were extracted (study design, the setting of the target population such as a hospital, sample size, age, sex, and population type such as physicians). We only included studies with a sample size of under 20 participants provided that the risk of bias was adequate [32].

We analyzed studies in terms of objectives, outcomes, and assessments as well as types of DHT. The quality criteria of assessments and information regarding the application of eye tracking as well as outcomes assessed via eye tracking were extracted. Differing from our protocol. we did not assess data on overall MWL in studies in addition to MWL levels related to DHT because the studies did not contain this information.

All included studies were evaluated with regard to their risk of bias.

A textual narrative synthesis of all included studies was made and comparable findings were synthesized. In addition, a descriptive analysis of eye tracking measures was extracted.

Registration and Protocol

In the ongoing process, we had to perform a few amendments.

Contrary to what was defined in the protocol to this review, research questions 1.1 and 1.2 were not substituted to this final paper [32]. Deviating from the protocol's attempt, we decided to use a different assessment tool to evaluate the risk of bias (QualSyst, [31]). In contrast to our protocol, we also included studies with a sample size under 20 participants under the condition that their risk of bias was adequate. Deviating from our protocol. we did not assess data on overall MWL in studies as well as MWL levels related to DHT because the studies did not contain this information.

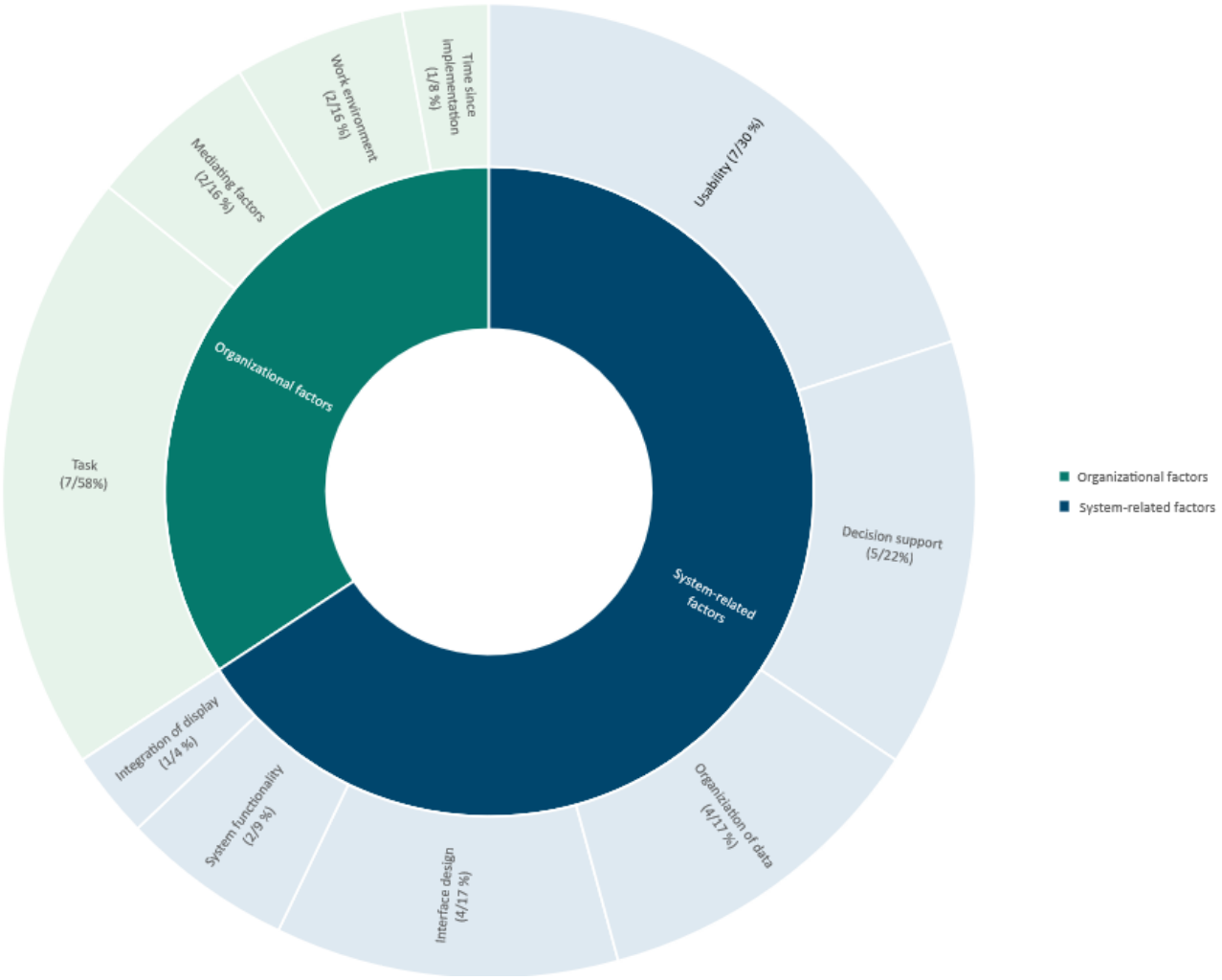
Results

Search Strategy

The database search resulted in 7952 hits. Additional searches in the bibliographies of the identified publications and through discussions with experts yielded 152 more search results (N=8104). After removal of duplicates**,** 6122 (75.54%) publications remained in the review process. On the basis of the title and abstract screening, 6003 (74.07%) publications were excluded. Of the remaining 117 (1.4%) that were included in the full-text analysis, 72 (62%) were excluded for the following reasons: another concept of stress than defined in our paper (eg, technostress) was used, DHT was not part of the study, the study outcome was not workload, the paper was not

an original work, the scope of the paper was alert-related workload, the population consisted of patients, it was a scoping or narrative review, there was no health care setting, or the full text was not available. In total, 46 (0.6%) studies were included in the qualitative synthesis and assessed for risk of bias. Of these, 17 (37%) studies were excluded because of their high risk of bias. The systematic search and the search strategy that followed resulted in 25 included studies (Figure 3).

Figure 3. Contributing factors to mental workload related to digital health technologies grouped into system related and organizational factors. The categories are not disjunct, meaning that two categories may have been selected for one study. The categories are not mutually exclusive either.



Risk of Bias Assessment

In total, 17 (37%) studies had a high risk of bias and were therefore excluded from the review because of scores <0.5.

A total 15 (33%) studies had scores between 0.5 and 0.65 and were therefore considered to have a moderate risk of bias. Furthermore, 10 (22%) studies had a low risk of bias (as shown in Multimedia Appendix 3; interrater agreement on scoring was r=0.91; P=.01).

Discrepancies in scoring generally resulted in different scores for item 1 (objectives) or 7 (blinding).

Main Characteristics of the Included Studies

The main characteristics of the included studies are displayed in Table 1. Most studies were published between the years 2010 and 2022 [33-54]. Only 2 studies were published between the years 2002 and 2009 [55,56]. Most studies were conducted and published in the United States [33,35,37-40,42-50,52,53,55,57].

Most studies were carried out in laboratory or simulation settings [33,35,36,38-43,46-50,52,55,56], a few were done in field settings [37,45,54,57], and some were conducted only on the web [34,44,51].

A total of 10 studies were quasi-experimental [33,36,40,41,46,47,49,50,55,57], 8 were cross-sectional [34,37,39,44,48,51,54,56], 2 were observational [38,53], 1 was a longitudinal design study [45], and 4 were RCTs [35,42,43,52].

The included participants consisted of physicians (14 studies [33-35,37-39,42-44,46-52,54,56,57]), nurses (4 studies [40,45,53,55]), and medical or nursing students (1 study [41]) as well as mixed populations out of these 3 groups (6 studies [36,37,42,47,50,57]). The sample size in most included studies ranged from 10 to 50 participants [33,35,36,38-40,42,43,46-50,52-57], 1 study ranged from 50 to 100 participants [34], and 5 studies included >100 participants [37,41,44,45,51]. Furthermore, 16 studies reported times of experience with DHT [33,34,36,40,42,43,45,48,49,51-53,55-57].

Table 1. Main characteristics of the included studies, including the display of sample statistics, setting, study design, and descriptive information about the included studies.

Author	Country	Setting	Study design	Sample size, n	Age (years), mean (SD) or median (IQR)	Sex, n (%)	Experience with DHT[a] (years)	Occupation
Ahmed et al [33], 2011	United States	L[b]	QS[c]	20	NR[d]	NR	>1 year	P[e]
Ariza et al [34], 2015	United Kingdom	Wb[f]	CS[g]	67	NR	NR	6.7 years	P
Carayon et al [35], 2020	United States	L	E[h]	32	NR	Female 8 (25); male 24 (75)	NR	P
Currie et al [36], 2017	United Kingdom	L	QS	S[i] 37; N[j] 11	S 27.31;N 31.91 SD or range NR	NR	S 0 years; N 8.73 years	N; S
Dunn Lopez et al [57], 2021	United States	F[k]	QS	N 22; *P* 13	N 32.5 (20-66); *P* 45.3 (25-63)	N: female 19.8 (90), male 2.2 (10); P: female 5.98 (46), male 7.02 (54)	N 2.5 years; P 6.2 years	N; P
Grünloh et al [54], 2016	Sweden	F	CS	12	NR	Female 5 (42); male 7 (58)	14 (2-30) years	P
Holden et al [37], 2015	United States	F	CS	170	NR	Female ≥161 (>95) Males: <9 (<5%)	NR	N; P
Khairat et al [38], 2018	United States	L	O[l]	14	Resident: 18-34 years (6, 100%) Attending: 35-50 years (7, 87.5%) 51-69 years (1, 12.5%)	Female 7 (50); male 7 (48)	Residents 3 years; Attending >3 years	P
Khairat et al [39], 2019	United States	L	CS	25	33.2 (6.1) years	Female 13 (52); male 12 (48)	NR	P
Koch et al [40], 2012	United States	L	QS	12	31.5 (23-57)	Female 8 (66); male 4 (34)	Self-rated experts 9 years; self-rated novices 1 year	N
Lyell et al [41], 2018	Australia	L	QS	120	24.5 (2.99)	Female 55.2 (46.7); male 63.6 (53.3)	NR	S
Mazur et al [42], 2015	United States	L	E	29	NR	NR	WebCIS 0.5-3 years; Epic 0.5 years	P; S
Mazur et al [43], 2019	United States	L	E	38	NR	Female 25 (66); male 13 (34)	Residents 36 years; fellows 2 years	P
Melnick et al [44], 2020	United States	Wb	CS	848	53 (28-84)	Female 509 (58.1); male 353 (40.6)	NR	P
Moreland et al [45], 2012	United States	F	Lo[m]	719	38.5 (11.2)	Female 650 (90.9); male 69 (9.1)	Participants self-rated “comfort with system” and sorted by group (n) Novice 41; knowledge of basics 288; experts 390	N
Mosaly et al [47], 2018	United States	L	QS	17	NR	NR	NR	P
Mosaly et al [46], 2019	United States	L	QS	38	NR	Female n (63); male n (27)	NR	P; S
Pollack et al [48], 2020	United States	L	CS	29	43 (35-58)	Female n (48); male n (52)	11 (3-30) years	P

Author	Country	Setting	Study design	Sample size, n	Age (years), mean (SD) or median (IQR)	Sex, n (%)	Experience with DHT[a] (years)	Occupation
Richardson et al [49], 2019	United States	L	QS	32	39.29 (12.4)	Female n (50); male n (50)	Participants (n); Residents (minimum of 3 years experiece) 16; attending physicians (Training level) 16	P
Saleem et al [55], 2007	United States	L	QS	16	NR	NR	None	N
Sampson et al [50], 2019	United States	L	QS	35	34.2 (25-59)	NR	NR	N; P
Shachak et al [56], 2009	Israel	L	CS	25	NR	Female n (56), male n (44)	6.8 years	P
Shah et al [51], 2016	United Kingdom	Wb	CS	188	NR	Female n (63.3); male n (36.7)	3-6 months: 54 (n) 6 months – 1 year: 51 (n) >1 year: 83 (n)	P
Wanderer et al [52], 2011	United States	L	E	20	NR	NR	Residents 10 years; attending physicians 10 years	P
Yen et al [53], 2020	United States	F	O	7	30 (6)	Female 6 (86); male 1 (14)	NR	N

[a]DHT: digital health technology.

[b]L: labor.

[c]QS: quasi-experimental.

[d]NR: not reported.

[e]P: physician.

[f]Wb: web-based.

[g]CS: cross-sectional.

[h]E: experimental.

[i]S: student.

[j]N: nurse.

[k]F: field.

[l]O: observational.

[m]Lo: longitudinal.

The included studies did not apply a homogenous definition approach for MWL: 13 studies did not provide a definition of their underlying concept at all [35,36,38,40-43,45,50,52-54,57], 2 studies applied a classic definition of MWL [37,51], 3 studies defined MWL as mental effort [34,46,47], 2 as information overload [33,39], and 5 studies applied a definition of cognitive load [41,44,48,49,56]. All the applied definition had a common base that could be summed up under the concept of MWL that we defined for inclusion.

The analyzed types of DHT were grouped into one of six categories as appropriate electronic health records or electronic medical records (EMRs), computerized decision support systems, information display or vital sign display, e-prescribing systems, anesthesia system, and computerized clinical reminders. More than half of the studies (13/25, 52%) analyzed electronic health records or EMRs.

Research Question 1: Contribution of DHT to the MWL of HCPs

Studies with various outcomes reflecting the association of DHT and MWL were included.

Overall, 20 (83%) of the included studies investigated the MWL related to DHT in general [33,38,40,53,54,56], 8.33% (2/25) compared MWL before and after redesign of DHT [51,52], and 12.5% (3/25) of the studies analyzed MWL before and after implementation of a new DHT [37,50,55]. A further 12.5% (3/25) of the studies compared MWL among different DHT or systems [34,35,40].

Furthermore, 33.33% (5/25) of the included studies investigated the relationship between the usability of the DHT and MWL [39,43,44,48,57], 16.67% (4/25) assessed MWL related to task demands and performance during the use of DHT [39,42,46,47], 8.33% (2/25) of the studies examined the influence of decision support on MWL [41,49], and 4% (1/25) examined other influences [36].

The included studies identified various factors of the systems that contributed to the MWL of HCPs. Some factors were rooted in the systems themselves; other factors were caused by influences and circumstances on an organizational level. We grouped the results by organizational and system-related factors (Figure 2).

Organizational Factors

A total of 8 studies identified the task to be performed by the use of the DHT as the relevant factor that contributes to an increasing MWL [34,36,37,41,42,47,50,54,56]. In all cases, the tasks did not fit the processes already implemented in the system.

Of these, 2 studies stated the overall workload in the working environment as the contributing factor [53,54]: the higher the general workload, the higher the MWL related to DHT.

Other relevant organizational factors that were identified by a study was the amount of time since implementation [45]: the longer a system was implemented, the lower the MWL, which initially increased significantly immediately after implementation.

In addition to direct influences, a study examined mediating factors and specifically identified gender and total hours worked. Women, as well as those who worked fewer hours, had a smaller increase in MWL from the DHT [44,57].

System Factors

In addition to organizational factors, most studies (23/25, 92%) identified factors based predominantly in the underlying system of the DHT.

A total of 4 studies cited weaknesses in the interface design as main factors for an increasing MWL of HCPs [39,48,50,51].

In addition to the interface design, 6 studies identified deficiencies in the usability as an influencing factor for increasing workload [39,43-45,50,51]. Studies refer to longer task completion times, higher error rates, a higher number of clicks, and differences in usability ratings between men and women (women contributed to higher rankings) [39]. There were also reports of less MWL because of automatically sorted and displayed test results in an electronic health record [43] and a significant correlation between MWL and usability [44].

A further 5 studies identified nonfunctioning decision support as a critical factor in increasing MWL [35,41,47,49,56], and 4 studies detected the organization of data and information as influencing factors [33,36,38,56].

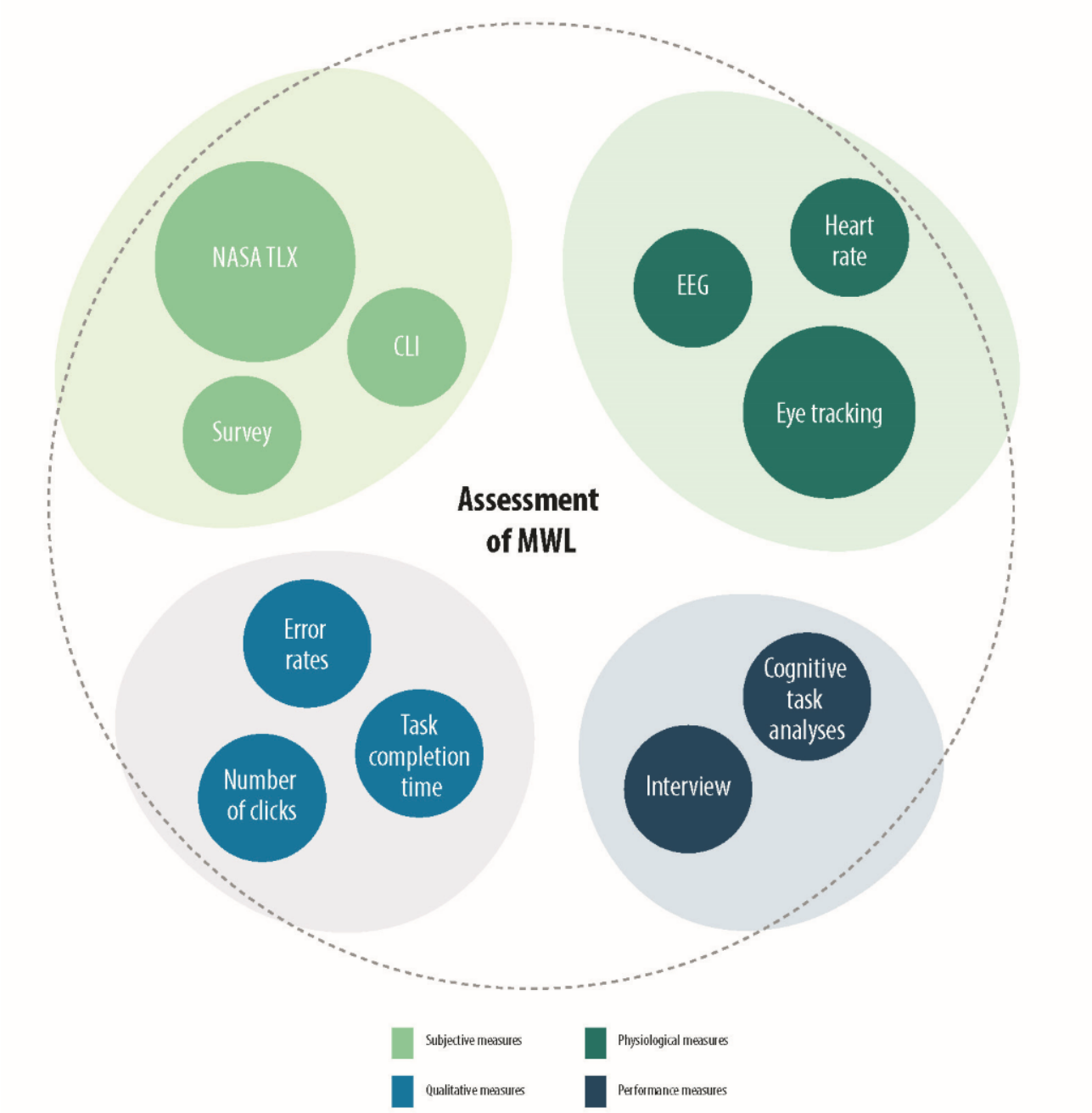
A study showed that integrated displays cause less MWL than nonintegrated traditional displays [40]. In addition to precisely identifiable factors, 2 studies indicated that high MWL is particularly because of system functionality of the system in itself [34,39].

Research Question 2: Assessment Methods of MWL Related to DHT in Health Care

Overview

All applied and identified assessment methods have been empirical. A total of 18 studies applied subjective methods [34-36,38,40-42,44,48-50,52,53,55,57], and 2 studies used performance measures [33,35]. Furthermore 5 studies used physiological methods [36,42,43,46,47], and all of them applied eye tracking techniques—either isolated or in combination with other measures [36,42,43,46,47]. In a study, an interview was conducted [54], and a study used the cognitive task analyses technique [56]. The identified measures are displayed in Figure 4.

Figure 4. Identified assessment methods grouped by assessment type. Most applied assessment type were subjective methods – NASA TLX was the assessment that was used in most studies. The size of the circles is proportional to the frequency of application in the studies.



Subjective Measures

National Aeronautics and Space Administration–Task Load Index or Raw–Task Load Index

A total of 52% (13/25) of the included studies applied the National Aeronautics and Space Administration (NASA)–Task Load Index (TLX) or an adapted form of the questionnaire such as the Raw-TLX to assess the MWL of HCPs in relation to DHT [34-36,38,40,42,44,48-50,52,53,55,57].

Of these, 3 (12%) studies adapted the NASA-TLX in form of the Raw-TLX based on numerous trials [34,35,51].

The NASA-TLX is a very commonly applied subjective assessment method to assess the MWL related to a specific task. The NASA-TLX has been applied mostly for questions of interface design and evaluation [58] and is often combined with other applied measures such as performance measures [58].

The questionnaire consists of six scales that each represent 1 dimension: MWL, physical workload, temporal workload, effort, frustration, and performance [59].

The original form of the NASA-TLX provides a rating scale ranging from 0 to 100 and a weighting of the different values of the scales [59]. However, several studies could show that the weighting of the scales in particular has no degrading influence on the sensitivity of the scales [58]. Thus, this form of the questionnaire is called the Raw-TLX and is the most commonly used version along with the NASA-TLX itself [58]. Even a change in the Likert scale does not seem to lead to a strong modification of the sensitivity or the quality criteria [58]. The

psychometrics for both versions, the original NASA-TLX and the Raw-TLX, can be considered good [60,61].

Cognitive Load Inventory

The cognitive load inventory was applied by a study and can be defined as a subjective cognitive load measurement tool [62]. Leppink et al [62] developed a 10-item questionnaire, rated on a 10-point Likert scale with the dimensions of intrinsic, extraneous, and germane load. The development of this scale was based on the cognitive load theory [63]. Previous research shows that psychometrics for this scale can be considered good [62].

Self-developed Surveys

A study used a self-developed survey that consisted of items for external (3 items) and internal (2 items) MWL. The Cronbach α for both scales was average to good [37].

Another study analyzed nurse workload by using 2 self-developed items that were rated on a 10-point Likert scale. Content validity (0.92) and internal consistency can be considered good (Cronbach α=.89-.95) [45].

Physiological Measures

Electroencephalography

Mazur et al [42] measured cognitive workload derived from electroencephalography. They processed the data by applying the ABM's algorithm that automatically calculates the index of cognitive workload.

Previous research shows that specific features of brain activity are good indicators for MWL; for example, theta activity increases with increasing mental effort [64].

Accuracy levels of electroencephalography measures can be classified as average (approximately 60%) [65].

Eye Tracking

A total of 5 studies applied eye tracking to measure the MWL related to DHT (displayed in Table 2). Furthermore, 3 studies assessed the blink rate of participants as an indicator for MWL [43,46,47], 2 studies detected pupil dilations [42,46], 1 study assessed fixation frequency and visit frequency [36], and 1 study applied the measure of task evoked pupillary response [47] None of these studies reported quality criteria for their assessment.

Table 2. Display of assessments of mental workload (MWL) via eye tracking.[a]

Study	Measures	Measure combination	Outcomes assessed
Currie et al [36], 2018	Visit frequency; fixation frequency	Questionnaire (NASA-TLX[b])	Automatic prediction of performance of nurses and interpretation of vital monitors
Mazur et al [42], 2016	Pupil dilations	Questionnaire (NASA-TLX); electroencephalography	Performance (error count and task completion time)
Mazur et al [43], 2019	Blink rate	N/A[c]	Mental and physical workload, performance, and fatigue
Mosaly et al [46], 2019	Blink rate; pupil dilations	N/A	Mental effort and performance
Mosaly et al [47], 2018	Blink rate; task evoked pupillary response	N/A	Mental effort and performance

[a]The most frequently applied measure was pupil dilation. The outcomes assessed varied across studies.

[b]TLX: Task Load Index.

[c]N/A: Not applicable.

Heart Rate

A study used a wearable heart rate monitor to detect heart rate changes as indicators of nurses' workload levels. The device assessed biometric signals continuously with time stamps. No psychometric values were given [53].

Performance Measures

A total of 2 studies applied performance measures as detection methods for MWL. Both studies did not apply these as stand-alone assessments; they combined the assessments with questionnaires. The response time, error rate, and number of clicks were measured.

Ahmed et al [33] registered the time to task completion (in seconds). Completion of tasks on a standard EMR in comparison to a redeveloped one took twice as long.

Ahmed et al [33] also counted the number of errors. They identified 4 times as many errors per participant when using the standard EMR than when using a redesigned user interface.

Carayon et al [35] assessed the number of clicks and task completion time and correlated these using the measures of NASA-TLX. Physicians were faster and interacted with lesser interface elements for a clinical decision support system when compared with the standard system.

Qualitative Measures

Shachak et al [56] applied a cognitive task analyses using semistructured interviews as well as field observations to assess MWL related to EMRs. The interview is adapted from the study by Militello and Hutton and asks for characteristics of the system that require difficult cognitive skills, errors, and special attention. Physicians reported a reduced MWL when EMR systems were used

Quality Criteria of Applied Methods

Overall, 68% (17/25) of the included studies did not report any quality criteria or measure. Some referred to reliability scores cited from previous research.

Furthermore, 5 (20%) studies reported measures of reliability (Cronbach α). Carayon et al [35], Lyell et al [41], and Moreland et al [45] reported a Cronbach α between.8 and .9.

Holden et al [37] and Shah and Peikari [51] reported a Cronbach α between.7 and .8.

In addition to Cronbach α, Moreland et al [45] reported a high content validity (0.9). A total of 2 (8%) studies reported quality criteria but only partly or not adequately [40,57].

Approach Toward the Most Applied Combination or Gold Standard

The combination of setting and applied measure that was detected in most cases was a laboratory setting combined with a subjective measurement method. Further, it can be identified that the outcome relationship between MWL and usability related to DHT measured by subjective methods or performance measures in the laboratory was established in most cases. Other frequently applied combinations were subjective method, MWL related to DHT, decision support, usability, system comparison, or other as well as physiological measures combined with task demands or other, in the laboratory. The results of the combinations of settings, assessments, and outcomes are displayed in Multimedia Appendix 4.

Discussion

Although several measures are applied frequently in the assessment of MWL in varied areas, the use of these methods may be limited by shortcomings in terms of knowledge about their correct and valid application in the field of human-technology interaction in health care. Therefore, our review had 2 separate but related objectives as described in the following sections.

Principal Findings

This systematic review investigated 25 studies that applied various measurement methods to assess the MWL related to DHT. The aim of the review was to show which factors of DHT contribute to a high MWL for HCPs in health care settings. In addition, the review was intended to identify methods that are currently used to measure MWL in health care. In this context, the role of eye tracking as a measurement method in particular was considered.

The following aspects can be considered the most relevant while summarizing the main results:

- First, the investigation showed that self-report subjective measurement methods (eg, the NASA-TLX), are the most frequently applied measures and can be considered the most prominent measure in MWL evaluation. Studies are most commonly conducted in laboratory settings. If physiological measures such as eye tracking are applied, they are combined with other measurement methods.
- Although a most frequent approach could be identified, it has to be stated that the methods used for the measurement of MWL related to DHT varied in their scope, methodology, outcomes, and evidence level as well as results concerning the MWL created by DHT.
- The risk of bias assessment revealed severe deficiencies in most studies because of methodological issues, inadequate sample sizes and statistical power, and poor study designs as well as deficient conduction of studies.

In particular, the negative effect of DHT on MWL in health care was consistent across studies. At the same time, DHT could support HCPs, but it must fulfill different criteria to achieve this. In addition to the system-related factors, organizational issues contribute to the influence of DHT on high MWL.

Comparison With Prior Work

Consistent with previous reviews, we identified the application of subjective measurement methods to be the most frequently used approach for the assessment of the MWL. [66]. Although we were able to identify a most frequently applied method, one of the main findings of this review was the heterogeneity of applied assessments, which is also in line with previous analyses [20,66]. Some studies used a combination of methods; for example, eye tracking and NASA-TLX. Reviews that investigate methods to measure MWL usually focus on 1 type of method, such as physiological measures [17,18], or a specific field of application (eg, driving distraction) [62]. The health care domain—although it can be seen as a safety-critical environment—was not the focus of these reviews. Charles and Nixon [21] included 58 studies in their review, none of which addressed MWL in health care. First, while other studies focused on nonhealth care domains, our review revealed methodological shortcomings in the health care area.

Second, our idea was to provide a holistic review of methods being used for the application of DHT in health care.

Previous reviews also checked for combined measure assessments; in line with our findings, Charles and Nixon [21] and Tao et al [20] found several studies that combined physiological measures and the NASA-TLX.

In contrast with our findings, Charles and Nixon [21] found many studies reporting quality criteria such as sensitivity and validity, also for physiological measures. However, they found differences for validity and sensitivity of measures comparing field and laboratory settings. This finding corresponds to the findings of Tao et al [20] and also partially to our findings.

Kabilmiharbi et al [22] reviewed studies concerning multiple driving distractions. In contrast to health care settings, MWL assessment during driving is mainly conducted via physiological or performance measures [63]. In line with our results, NASA-TLX was the most commonly used subjective assessment.

We identified 4 different eye tracking measures applied in the studies included in our review (fixation frequency, blink rate, pupil dilation, and visit frequency). Tao et al [20] identified blink rate, pupil diameter, and fixation duration as correlates of

MWL, but—in contrast with our results—identified additional eye tracking measures that were relevant.

Besides a strong heterogeneity, a rather homogeneous approach with regard to the setting was revealed. This is equivalent to findings of Tao et al [20]. Most studies were performed in the laboratory. Outcomes differed marginally but were still differentiated for more discriminative analysis.

Factors contributing to MWL in health care can be identified as occupational or individual. Occupational factors can be level of education, type of working unit (eg, intensive care unit), work shifts, and number of patients under care [64]. Studies from other domains show, for example, an enhancement in situation complexity, task-related and individual factors as well as organizational factors such as time pressure as possible predictors of MWL [65]. However, none of the studies mentioned in this section explicitly addresses the relationship between MWL and HIS/DHT.

Many studies also consider MWL as a starting point for further consequences on the performance of the HCPs, for example, a hazard to patient safety or job satisfaction (66), rather than the factors contributing to a high MWL.

Strengths and Limitations

This review has some limitations with respect to the included studies.

First, because of the heterogeneity of the assessment methods, analyses, and study designs of the included studies as well as their methodological quality, a meta-analysis could not be conducted.

Second, many studies performed retrospective measurements of MWL that did not allow for causal conclusions in the results. The restriction of causality is further limited by nonreported quality criteria.

Third, the results as well as the review itself are further limited by the search process. Part of the results are aspects of factors that contribute to MWL related to DHT. These aspects were not explicitly searched for in the literature examination. It can therefore be assumed that not all relevant studies concerning these factors have been included. The search process can also be considered to be limited in the sense that it became apparent during the review process that many authors integrate the constructs of mental or cognitive workload into other constructs or refer to concepts similar to these. Other constructs that may follow a similar definition, such as mental effort, were not considered in this search. It can therefore be assumed that these studies were not included in the review.

The definition of the MWL construct was not consistent across the studies examined. In addition to MWL, stress, cognitive load, fatigue, and mental effort, and other similar concepts have been grouped under the term information overload and limited workload capacity resulting from perceptual load. However, other studies have developed their own concepts (eg, stress related to information systems) that mean slightly different things but include parts of the definition of MWL. Our results are limited in terms of not including these studies as they also included aspects of stress (eg, acceptance) that do not refer to the MWL classification that was relevant for our paper.

However, in order to develop a gold standard for measuring MWL in health care settings, it seems highly relevant to precisely define the construct. Identifying studies referring to a selective definition of MWL was therefore particularly challenging for this review. Because of the strong heterogeneity of the research field, we cannot eliminate the possibility that some studies were not included, which were not identified by our search terms because of variations in construct naming.

The combination of the different approaches toward the assessment of MWL also showed strong heterogeneity. Some of the methods—especially the physiological ones—require extensive preparation and equipment and are very time-consuming, particularly in their evaluation. Thus, not every method can be considered suitable for every setting (eg, in a clinical setting).

The approach of analysis in the laboratory seems understandable on the one hand, because content validity and reliability are easy to achieve. On the other hand, the small number of field studies ensures that results cannot be transferred to other settings easily (external validity) and that various bias effects at least partly due to presumably weak quality of the study implementation also led to erroneous results. This also applies to the generalizability across populations; therefore, studies referring to MWL of patients were not included.

The applied quality criteria assessment revealed shortcomings in methodological quality across many studies. There was only a small amount of studies with a quality rating of >65% (10 studies [37,39,41-44,49,50,54-56]). However, a possible explanation for such a low rate might be that many of the remaining studies could be regarded as first or exploratory approaches.

Most studies did not report quality criteria such as content validity or reliability. Reliability indicates the degree to which an assessment can differ between high and low workloads [67]. Content validity refers to the degree to which an assessment reflects all aspects of MWL [67]. Studies that reported reliability measures reported acceptable to high levels of internal consistency of the assessments. Studies that reported content validity reported moderate levels of internal consistency of the assessments. To develop a gold standard in the assessment of MWL in health care, the reporting of quality criteria as indications for the quality of a measurement method is essential.

Studies that were not published in full text or in English were excluded; consequently, additional information on measurement properties and descriptions of methods for assessing masticatory performance that may have potentially affected the level of evidence might have been missed.

All included papers were published in the period between 2002 and 2022; the literature search was limited to papers with publication years between 2000 and 2022.

We detected an increase in the 2010s that could give a hint regarding the increasing interest in the topic during this time. On the other hand, the term MWL, as already described, was

not defined in as much detail as it should have been. Therefore, the detected increase could have also been produced by more specific definitions in the last years.

In addition, it is possible that we did not find all relevant articles, despite having thoroughly defined which terms to include and having conducted a systematic search using Medical Subject Heading terms

Future Directions

Our results show a very heterogenic approach toward the assessment of MWL related to DHT in health care settings. Although the assessments are heterogeneous, it can be assumed that there are 2 groups of contributing factors to MWL related to DHT, factors rooted in the system itself and organizational factors such as the task for which the system is being used.

When it comes to implementing or applying already implemented DHT in health care, these factors should be considered holistically.

The following steps should be taken for implementing and developing a gold standard and conducting future research in this field of study:

1. Conducting well-developed studies that take into account quality criteria and adequate sample sizes as well as effect size and power calculation. Future research is warranted to include HCPs with more diverse backgrounds (eg, differentiated by previous experience with DHT) and to have adequate statistical power for testing.
2. Reviewing MWL studies in related fields, such as power plants or aviation research.
3. Identifying methods that apply most to the research question being posed (eg, what is the amount of MWL of an intensive care unit nurse during a shift when switching between the EMR system and vital signs monitors), which would probably lead to a dynamic approach assessed by a dynamic assessment method such as eye tracking.

Future research is required to further investigate the relations between factors that might be contributing to MWL while using a DHT and MWL in general. Our results show a first step forward for grouping these factors. However, further primary research and review work is necessary for the development of a theoretical framework.

Conclusions

Our review of 25 papers shows a diverse assessment approach toward the MWL of HCPs related to DHT as well as 2 groups of relevant contributing factors to MWL. The most frequently applied method has been the NASA-TLX (subjective measurement approach) in laboratory settings. The contributing factors can be divided into system-related factors and organizational factors.

Our results show a few new approaches being used for assessing MWL in relation to systems in a valid, reliable and practical way; eye tracking could be one of these measurement techniques.

Although methodological biases were identified, we recommend further research concentrating on adequate assessments of MWL of HCPs for relevant settings. We would also like to recommend the evaluation of quality criteria.

Authors' Contributions

LK and BB conceived this study and screened the literature. LK drafted the topic of the study, wrote the manuscript and supervised the editing of the manuscript. BB, ML, and RR reviewed the manuscript. All authors approved this version to be published and agreed to be accountable for all aspects of the work with regard to ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Block chain of keywords that were used to create the search terms.
[DOC File , 31 KB - jmir_v24i10e40946_app1.doc]

Multimedia Appendix 2
Search string and results.
[PDF File (Adobe PDF File), 194 KB - jmir_v24i10e40946_app2.pdf]

Multimedia Appendix 3
Display of assessment scores of Qualsyst tool of both reviewers (rater LK and BB).
[XLSX File (Microsoft Excel File), 9 KB - jmir_v24i10e40946_app3.xlsx]

Multimedia Appendix 4

Tabular display of the descriptive results relating to a combination of specific outcomes of the review: The table displays results of single categories (e.g. subjective method) on the one hand, and on the other hand combined results of several categories (e.g. applied method and outcome of study).
[XLS File (Microsoft Excel File), 60 KB - jmir_v24i10e40946_app4.xls]

References

1. Kostera T, Thranberend T. Spotlight Gesundheit: #SmartHealthSystems: Digitalisierung braucht effektive Strategie, politische Führung und eine koordinierende Institution. Bertelsmann Stiftung. Gütersloh, Germany: Bertelsmann Stiftung; 2018. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/spotlight-gesundheit-smarthealthsystems [accessed 2022-10-19]
2. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern Med 2013 Apr 22;173(8):702-704 [FREE Full text] [doi: 10.1001/2013.jamainternmed.61] [Medline: 23460235]
3. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
4. Trayambak T, Singh AL, Singh IL. Information technology-induced stress and human performance: a critical review. J Indian Acad Appl Psychol 2008 Jul;34(2):241-249.
5. Dang YM, Zhang YG, Brown SA, Chen H. Examining the impacts of mental workload and task-technology fit on user acceptance of the social media search system. Inf Syst Front 2020;22(3):697-718. [doi: 10.1007/s10796-018-9879-y]
6. Proctor RW, Van Zandt T. Human Factors in Simple and Complex Systems. 2nd edition. Boca Raton, FL, USA: CRC Press; 2008.
7. Smith KT. Observations and issues in the application of cognitive workload modelling for decision making in complex time-critical environments. In: Proceedings of the 1st International Symposium on Human Mental Workload: Models and Applications. 2017 Presented at: H-WORKLOAD '17; June 28-30, 2017; Dublin, Ireland p. 77-89. [doi: 10.1007/978-3-319-61061-0_5]
8. Wickens CD, Hollands JG, Banbury S, Parasuraman R. Engineering Psychology and Human Performance. 4th edition. London, UK: Routledge; 2016.
9. López-Núñez MI, Rubio-Valdehita S, Diaz-Ramiro EM, Aparicio-García ME. Psychological capital, workload, and burnout: what's new? The impact of personal accomplishment to promote sustainable working conditions. Sustainability 2020 Oct 01;12(19):8124. [doi: 10.3390/su12198124]
10. Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of electronic health record use with physician fatigue and efficiency. JAMA Netw Open 2020 Jun 01;3(6):e207385 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.7385] [Medline: 32515799]
11. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clin Proc 2020 Mar;95(3):476-487 [FREE Full text] [doi: 10.1016/j.mayocp.2019.09.024] [Medline: 31735343]
12. Heinke W, Dunkel P, Brähler E, Nübling M, Riedel-Heller S, Kaisers UX. Burnout in anesthesiology and intensive care : is there a problem in Germany? Anaesthesist 2011 Dec;60(12):1109-1118. [doi: 10.1007/s00101-011-1947-3] [Medline: 22071874]
13. Lysaght RJ, Hill SG, Dick AO, Plamondon BD, Linton PM. Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies. Springfield, VA, USA: United States Army Research Institute for the Behavioral and Social Sciences; 1989.
14. Wieland-Eckelmann R. Kognition, Emotion und psychische Beanspruchung: Theoretische und empirische Studien zu informationsverarbeitenden Tätigkeiten. Göttingen, Germany: Hogrefe Publishing; 1990.
15. Eggemeier FT, Wilson GF, Kramer AF, Damos DL. Workload assessment in multi-task environments. In: Damos DL, editor. Multiple-Task Performance. Boca Raton, FL, USA: CRC Press; 1991:217-278.
16. Gopher D, Donchin E. Workload: an examination of the concept. In: Boff R, Kaufman L, Thomas JP, editors. Handbook of Perception and Human Performance, Vol. 2. Cognitive Processes and Performance. Oxford, UK: John Wiley & Sons; 1986:1-49.
17. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 1908 Nov;18(5):459-482 [FREE Full text] [doi: 10.1002/cne.920180503]
18. Bruggen A. An empirical investigation of the relationship between workload and performance. Manag Decis 2015;53(10):2377-2389. [doi: 10.1108/md-02-2015-0063]
19. Tsang PS, Wilson GF, Salvendy G. Mental workload. In: Salvendy G, Karwowski W, editors. Handbook of Human Factors and Ergonomics. Hoboken, NJ, USA: Wiley; 1997:417-449.
20. Tao D, Tan H, Wang H, Zhang X, Qu X, Zhang T. A systematic review of physiological measures of mental workload. Int J Environ Res Public Health 2019 Jul 30;16(15):2716 [FREE Full text] [doi: 10.3390/ijerph16152716] [Medline: 31366058]
21. Charles RL, Nixon J. Measuring mental workload using physiological measures: a systematic review. Appl Ergon 2019 Jan;74:221-232. [doi: 10.1016/j.apergo.2018.08.028] [Medline: 30487103]

22. Körber M, Schmid K, Drexler H, Kiesel J. Subjective workload, job satisfaction, and work-life-balance of physicians and nurses in a municipal hospital in a rural area compared to an urban university hospital. Gesundheitswesen 2018 May;80(5):444-452. [doi: 10.1055/s-0042-121596] [Medline: 28201830]
23. Mark G, Smith AP. Occupational stress, job characteristics, coping, and the mental health of nurses. Br J Health Psychol 2012 Sep;17(3):505-521. [doi: 10.1111/j.2044-8287.2011.02051.x] [Medline: 22107162]
24. Baye Y, Demeke T, Birhan N, Semahegn A, Birhanu S. Nurses' work-related stress and associated factors in governmental hospitals in Harar, Eastern Ethiopia: a cross-sectional study. PLoS One 2020 Aug 3;15(8):e0236782 [FREE Full text] [doi: 10.1371/journal.pone.0236782] [Medline: 32745142]
25. Molina-Praena J, Ramirez-Baena L, Gómez-Urquiza JL, Cañadas GR, De la Fuente EI, Cañadas-De la Fuente GA. Levels of burnout and risk factors in medical area nurses: a meta-analytic study. Int J Environ Res Public Health 2018 Dec 10;15(12):2800 [FREE Full text] [doi: 10.3390/ijerph15122800] [Medline: 30544672]
26. Sturm H, Rieger MA, Martus P, Ueding E, Wagner A, Holderried M, WorkSafeMed Consortium. Do perceived working conditions and patient safety culture correlate with objective workload and patient outcomes: a cross-sectional explorative study from a German university hospital. PLoS One 2019 Jan 4;14(1):e0209487 [FREE Full text] [doi: 10.1371/journal.pone.0209487] [Medline: 30608945]
27. Marschall J, Hildrebrandt S, Kleinlercher KM, Nolting HD. Gesundheitsreport 2020: Stress in der modernen Arbeitswelt. Sonderanalyse: Digitalisierung und Homeoffice in der Corona-Krise. Heidelberg, Germany: medhochzwei Verlag; 2020.
28. 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]
29. Rogers H, Madathil KC, Agnisarman S, Narasimha S, Ashok A, Nair A, et al. A systematic review of the implementation challenges of telemedicine systems in ambulances. Telemed J E Health 2017 Sep;23(9):707-717. [doi: 10.1089/tmj.2016.0248] [Medline: 28294704]
30. 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]
31. Kmet LM, Lee RC, Cook LS. Standard Quality Assessment Criteria for Evaluating Primary Research Papers From a Variety of Fields. Edmonton, Canada: Alberta Heritage Foundation for Medical Research; 2004.
32. Kremer L, Lipprandt M, Röhrig R, Breil B. Examining the mental workload associated with digital health technologies in health care: protocol for a systematic review focusing on assessment methods. JMIR Res Protoc 2021 Aug 03;10(8):e29126 [FREE Full text] [doi: 10.2196/29126] [Medline: 34342590]
33. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit Care Med 2011 Jul;39(7):1626-1634. [doi: 10.1097/CCM.0b013e31821858a0] [Medline: 21478739]
34. Ariza F, Kalra D, Potts HW. How do clinical information systems affect the cognitive demands of general practitioners?: usability study with a focus on cognitive workload. J Innov Health Inform 2015 Nov 20;22(4):379-390 [FREE Full text] [doi: 10.14236/jhi.v22i4.85] [Medline: 26855272]
35. Carayon P, Hoonakker P, Hundt AS, Salwei M, Wiegmann D, Brown RL, et al. Application of human factors to improve usability of clinical decision support for diagnostic decision-making: a scenario-based simulation study. BMJ Qual Saf 2020 Apr;29(4):329-340 [FREE Full text] [doi: 10.1136/bmjqs-2019-009857] [Medline: 31776197]
36. Currie J, Bond RR, McCullagh P, Black P, Finlay DD, Peace A. Eye tracking the visual attention of nurses interpreting simulated vital signs scenarios: mining metrics to discriminate between performance level. IEEE Trans Human Mach Syst 2018 Apr;48(2):113-124. [doi: 10.1109/THMS.2017.2754880]
37. Holden RJ, Brown RL, Scanlon MC, Rivera AJ, Karsh BT. Micro- and macroergonomic changes in mental workload and medication safety following the implementation of new health IT. Int J Ind Ergon 2015 Sep;49:131-143. [doi: 10.1016/j.ergon.2014.04.003]
38. Khairat S, Burke G, Archambault H, Schwartz T, Larson J, Ratwani R. Perceived burden of EHRs on physicians at different stages of their career. Appl Clin Inform 2018 Apr;9(2):336-347 [FREE Full text] [doi: 10.1055/s-0038-1648222] [Medline: 29768634]
39. Khairat S, Coleman C, Ottmar P, Bice T, Koppel R, Carson SS. Physicians' gender and their use of electronic health records: findings from a mixed-methods usability study. J Am Med Inform Assoc 2019 Dec 01;26(12):1505-1514 [FREE Full text] [doi: 10.1093/jamia/ocz126] [Medline: 31504578]
40. Koch SH, Westenskow D, Weir C, Agutter J, Haar M, Görges M, et al. ICU nurses' evaluations of integrated information displays on user satisfaction and perceived mental workload. Stud Health Technol Inform 2012;180:383-387. [Medline: 22874217]
41. Lyell D, Magrabi F, Coiera E. The effect of cognitive load and task complexity on automation bias in electronic prescribing. Hum Factors 2018 Nov;60(7):1008-1021. [doi: 10.1177/0018720818781224] [Medline: 29939764]
42. Mazur LM, Mosaly PR, Moore C, Comitz E, Yu F, Falchook AD, et al. Toward a better understanding of task demands, workload, and performance during physician-computer interactions. J Am Med Inform Assoc 2016 Nov;23(6):1113-1120 [FREE Full text] [doi: 10.1093/jamia/ocw016] [Medline: 27026617]

43. Mazur LM, Mosaly PR, Moore C, Marks L. Association of the usability of electronic health records with cognitive workload and performance levels among physicians. JAMA Netw Open 2019 Apr 05;2(4):e191709 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.1709] [Medline: 30951160]
44. Melnick ER, Harry E, Sinsky CA, Dyrbye LN, Wang H, Trockel MT, et al. Perceived electronic health record usability as a predictor of task load and burnout among US physicians: mediation analysis. J Med Internet Res 2020 Dec 22;22(12):e23382 [FREE Full text] [doi: 10.2196/23382] [Medline: 33289493]
45. Moreland PJ, Gallagher S, Bena JF, Morrison S, Albert NM. Nursing satisfaction with implementation of electronic medication administration record. Comput Inform Nurs 2012 Feb;30(2):97-103. [doi: 10.1097/NCN.0b013e318224b54e] [Medline: 21993338]
46. Mosaly PR, Guo H, Mazur L. Toward better understanding of task difficulty during physicians' interaction with electronic health record system (EHRs). Int J Human Comput Interact 2019 Feb 20;35(20):1883-1891. [doi: 10.1080/10447318.2019.1575081]
47. Mosaly PR, Mazur LM, Yu F, Guo H, Derek M, Laidlaw DH, et al. Relating task demand, mental effort and task difficulty with physicians' performance during interactions with electronic health records (EHRs). Int J Human Comput Interact 2018;34(5):467-475. [doi: 10.1080/10447318.2017.1365459]
48. Pollack AH, Pratt W. Association of health record visualizations with physicians' cognitive load when prioritizing hospitalized patients. JAMA Netw Open 2020 Jan 03;3(1):e1919301 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19301] [Medline: 31940040]
49. Richardson KM, Fouquet SD, Kerns E, McCulloh RJ. Impact of mobile device-based clinical decision support tool on guideline adherence and mental workload. Acad Pediatr 2019;19(7):828-834 [FREE Full text] [doi: 10.1016/j.acap.2019.03.001] [Medline: 30853573]
50. Sampson JB, Lee BH, Koka R, Chima AM, Jackson EV, Ogbuagu OO, et al. Human factors evaluation of the universal anaesthesia machine: assessing equipment with high-fidelity simulation prior to deployment in a resource-constrained environment. J Natl Med Assoc 2019 Oct;111(5):490-499. [doi: 10.1016/j.jnma.2019.03.005] [Medline: 31078287]
51. Shah MH, Peikari HR. Electronic prescribing usability: reduction of mental workload and prescribing errors among community physicians. Telemed J E Health 2016 Jan;22(1):36-44. [doi: 10.1089/tmj.2014.0246] [Medline: 26244488]
52. Wanderer JP, Rao AV, Rothwell SH, Ehrenfeld JM. Comparing two anesthesia information management system user interfaces: a usability evaluation. Can J Anaesth 2012 Nov;59(11):1023-1031. [doi: 10.1007/s12630-012-9771-z] [Medline: 23055030]
53. Yen PY, Pearl N, Jethro C, Cooney E, McNeil B, Chen L, et al. Nurses' stress associated with nursing activities and electronic health records: data triangulation from continuous stress monitoring, perceived workload, and a time motion study. AMIA Annu Symp Proc 2019 Mar 4;2019:952-961 [FREE Full text] [Medline: 32308892]
54. Grünloh C, Cajander Å, Myreteg G. "The record is our work tool!"-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]
55. Saleem JJ, Patterson ES, Militello L, Anders S, Falciglia M, Wissman JA, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. J Am Med Inform Assoc 2007;14(5):632-640 [FREE Full text] [doi: 10.1197/jamia.M2163] [Medline: 17600106]
56. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009 Mar;24(3):341-348 [FREE Full text] [doi: 10.1007/s11606-008-0892-6] [Medline: 19130148]
57. Dunn Lopez K, Chin CL, Leitão Azevedo RF, Kaushik V, Roy B, Schuh W, et al. Electronic health record usability and workload changes over time for provider and nursing staff following transition to new EHR. Appl Ergon 2021 May;93:103359. [doi: 10.1016/j.apergo.2021.103359] [Medline: 33556884]
58. Hart SG. Nasa-task load index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet 2006 Oct;50(9):904-908 [FREE Full text] [doi: 10.1177/154193120605000909]
59. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human Mental Workload. Amsterdam, The Netherlands: Elsevier; 1988:139-183.
60. Devos H, Gustafson K, Ahmadnezhad P, Liao K, Mahnken JD, Brooks WM, et al. Psychometric properties of NASA-TLX and index of cognitive activity as measures of cognitive workload in older adults. Brain Sci 2020 Dec 16;10(12):994 [FREE Full text] [doi: 10.3390/brainsci10120994] [Medline: 33339224]
61. Longo L. On the reliability, validity and sensitivity of three mental workload assessment techniques for the evaluation of instructional designs: a case study in a third-level course. In: Proceedings of the 10th International Conference on Computer Supported Education. 2018 Presented at: CSEDU '18; March 15-17, 2018; Funchal, Portugal p. 166-178. [doi: 10.5220/0006801801660178]
62. Leppink J, Paas F, Van der Vleuten CP, Van Gog T, Van Merriënboer JJ. Development of an instrument for measuring different types of cognitive load. Behav Res Methods 2013 Dec;45(4):1058-1072. [doi: 10.3758/s13428-013-0334-1] [Medline: 23572251]
63. Sweller J, Chandler P, Tierney P, Cooper M. Cognitive load as a factor in the structuring of technical material. J Exp Psychol 1990 Jun;119(2):176-192. [doi: 10.1037/0096-3445.119.2.176]

64. Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Brain Res Rev 1999 Apr;29(2-3):169-195. [doi: 10.1016/s0165-0173(98)00056-3] [Medline: 10209231]
65. So WK, Wong SW, Mak JN, Chan RH. An evaluation of mental workload with frontal EEG. PLoS One 2017 Apr 17;12(4):e0174949 [FREE Full text] [doi: 10.1371/journal.pone.0174949] [Medline: 28414729]
66. Thorpe A, Nesbitt K, Eidels A. A systematic review of empirical measures of workload capacity. ACM Trans Appl Percept 2020 Nov 25;17(3):1-26. [doi: 10.1145/3422869]
67. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge, UK: Cambridge University Press; 2011.

Abbreviations

DHT: digital health technology
EMR: electronic medical record
HCP: health care professional
HIS: health information system
MWL: mental workload
NASA: National Aeronautics and Space Administration
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
TLX: Task Load Index

Edited by G Eysenbach, T Leung; submitted 10.07.22; peer-reviewed by S Sarejloo, Z Galavi, Z Dai; comments to author 08.08.22; revised version received 22.08.22; accepted 07.09.22; published 28.10.22.

Please cite as:
Kremer L, Lipprandt M, Röhrig R, Breil B
Examining Mental Workload Relating to Digital Health Technologies in Health Care: Systematic Review
J Med Internet Res 2022;24(10):e40946
URL: https://www.jmir.org/2022/10/e40946
doi:10.2196/40946
PMID:36306159

Review

Interventions Including Smart Technology Compared With Face-to-face Physical Activity Interventions in Older Adults: Systematic Review and Meta-analysis

Cassandra D'Amore[1*], BSc (Hons); Julie C Reid[1*], PT, PhD; Matthew Chan[1], MScPT; Samuel Fan[1], MScPT; Amanda Huang[1], MScPT; Jonathan Louie[1], MScPT; Andy Tran[1], MScPT; Stephanie Chauvin[2], MSc; Marla K Beauchamp[1,3,4], PT, PhD

[1]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

[2]Department of Physical Therapy, University of Toronto, Toronto, ON, Canada

[3]Department of Medicine, McMaster University, Hamilton, ON, Canada

[4]Research Institute at St Joseph's Healthcare, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Marla K Beauchamp, PT, PhD
School of Rehabilitation Science
McMaster University
1400 Main Street West
Institute for Applied Health Sciences Building
Hamilton, ON, L8S 1C7
Canada
Phone: 1 905 525 9140 ext 21732
Fax: 1 905 524 0069
Email: beaucm1@mcmaster.ca

Abstract

Background: This is a systematic review of randomized controlled trials and a meta-analysis comparing smart technology with face-to-face physical activity (PA) interventions in community-dwelling older adults (mean age 60 years).

Objective: This study aims to determine the effect of interventions including smart technology components compared with face-to-face PA interventions on PA and physical function in older adults. The secondary outcomes are depression, anxiety, and health-related quality of life.

Methods: We searched MEDLINE, Embase, CINAHL, and AMED electronic databases from inception to February 2021. Two independent reviewers screened titles, abstracts, and full texts and performed data extraction and risk of bias assessments using the Cochrane risk of bias tool. The Grading of Recommendations Assessment, Development and Evaluation was used to evaluate the quality of the evidence. We provided a narrative synthesis on all included studies and, where possible, performed meta-analyses for similar outcomes.

Results: This review included 19 studies with a total of 3455 participants. Random effects meta-analyses showed that interventions with smart technology components resulted in improved step count (mean difference 1440 steps, 95% CI 500-2390) and total PA (standardized mean difference 0.17, 95% CI 0.02-0.32) compared with face-to-face alone. There was no difference between groups in terms of the measures of physical function. Smart technology alone did not show significant differences between groups in any outcome. The quality of the evidence was very low based on the Grading of Recommendations Assessment, Development and Evaluation criteria.

Conclusions: Interventions that include smart technology may improve daily step counts by an average of 1440 steps in community-dwelling older adults; however, the quality of the evidence was very low. Future studies are needed to improve the certainty of these results.

Trial Registration: PROSPERO CRD42020135232; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=135232

(J Med Internet Res 2022;24(10):e36134) doi:10.2196/36134

KEYWORDS

aging; exercise; mobile health; mHealth; wearables; mobile phone

Introduction

Background

In 2017, the global population of adults aged ≥60 years was 962 million, more than twice the number of older adults in 1980 [1]. By 2050, it is expected that the number of older adults will double, reaching nearly 2.1 billion [1]. As the population ages, delaying the onset of illness and disability and retaining physical function are top public health priorities [2]. Physical activity (PA), defined as any bodily movement produced by skeletal muscles that requires energy expenditure [3], is one way to achieve this. However, evidence suggests that 31% of the global population does not meet the recommended levels of PA [4,5], and inactivity has been identified as a leading risk factor for mortality, accounting for >5 million global deaths annually [6]. A recent umbrella review including 24 systematic reviews reported that older adults that are physically active have a lower risk of all-cause and cardiovascular mortality, breast and prostate cancer, fractures, disabilities with activities of daily living, functional limitations, risk of falling, cognitive decline, dementia, Alzheimer disease, and depression [7]. In 2018, the World Health Organization released their global action plan on PA to combat inactivity and improve health over the next decade [8].

In late 2019, SARS-CoV-2, which causes COVID-19 disease, emerged and quickly became an international health crisis, and in March 2020, the World Health Organization declared COVID-19 a global pandemic [9]. Since then, many countries have established strict public health measures to curb the spread of the disease, including social distancing and isolation. Although these measures have the benefit of minimizing viral transmission, which is critical for older adults who are at a higher risk for more severe illness [10,11], they have also exacerbated levels of physical inactivity. A systematic review including 66 studies with nearly 87,000 participants from 26 countries reported significant declines in PA during the lockdown owing to the COVID-19 pandemic [12]. Unfortunately, older adults are also at a higher risk for consequences of inactivity, such as frailty, sarcopenia, and chronic diseases, compared with their younger counterparts [13]. These data highlight an urgent need to evaluate alternative methods of improving PA levels. Fortunately, with advancements in technology, smart technology has become an increasingly relevant and studied tool for achieving health objectives [14,15]. Smart technology interventions may represent an ideal alternative to traditional face-to-face programs as they have the potential to overcome service delivery barriers such as limited access; inconvenience of travel; absenteeism from work and family [16]; and, now importantly, minimizing unnecessary exposure to COVID-19 for those who are most at risk. The role of smart technology in improving PA in older adults warrants further evaluation both now and for informing future directions of health care delivery.

Smart technology capitalizes on communication and information technologies (eg, internet and video calls) [14] and uses different mediums, such as computers and tablets, or mobile health, which includes smartphones, wearables (eg, FitBit), and mobile apps (eg, My Fitness Pal, Samsung Health, and Apple Health) [17]. Systematic reviews have demonstrated that smart technology interventions can improve PA levels, specifically steps per day and minutes per day of moderate to vigorous PA in generally healthy older adult populations (mean age ≥55 years) [18-21]. However, there are still unanswered questions and a need for more and better evidence. For example, existing reviews have examined only specific types of smart technology [20] or included only digital PA estimates without considering participant-oriented outcomes [19]. Importantly, existing systematic reviews have not compared smart technology PA interventions with more traditional modes of PA intervention delivery (ie, face-to-face) [15,20-23]. This comparison is essential for determining whether interventions that include smart technology components are more, less, or as effective as face-to-face alone interventions. Therefore, the purpose of this review is to determine the effects of PA interventions that use smart technology compared with face-to-face PA interventions on PA and physical function in community-dwelling older adults.

Review Question

This systematic review will answer whether the PA interventions that use smart technology are more, less, or as effective as face-to-face alone interventions for increasing PA and function in older adults. The secondary questions were as follows: (1) What are the effects of smart technology PA interventions on secondary outcomes, including health-related quality of life (HRQoL), anxiety, and depression? and (2) Does the effectiveness of smart technology interventions differ by type of PA or by the type of smart technology used (eg, wearable vs mobile app)?

Methods

Overview

This systematic review was conducted in accordance with a peer-reviewed protocol [24] registered in PROSPERO (CRD42020135232) and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines [25]. The full protocol has been published elsewhere [24].

Data Source and Searches

A comprehensive search of the MEDLINE, Embase, CINAHL, and AMED databases from inception to February 2021 was conducted after consultation with a health research librarian [26]. The full search strategies are available in Multimedia Appendix 1 [27-45]; common Medical Subject Headings across databases included age, technology, physical fitness, with keywords to capture all types of PA and smart technology.

Reference lists of included studies were hand searched to identify additional relevant studies.

Study Selection

Overview

Two independent reviewers completed screening for both the titles and abstracts and full-text articles using the web-based referencing software system Covidence (Veritas Health Innovation). Disagreements were resolved by consensus or arbitration by a third reviewer as necessary.

Inclusion Criteria

We included studies that met the following criteria: (1) community-dwelling older adults with a mean age of ≥60 years [46], (2) interventions that promoted PA using smart technology, (3) face-to-face interventions in comparator groups, (4) a primary outcome measure of PA or physical function, and (5) randomized controlled trials (RCTs) published in English in a peer-reviewed journal.

Exclusion Criteria

We excluded studies that evaluated participants admitted to an inpatient unit in a hospital or long-term care home, interventions that only used audio phone calls (ie, with no video or SMS text messaging equivalent to the use of a landline), video games, or virtual reality. Studies that used a quasi-experimental design were also excluded.

Data Extraction and Risk of Bias

Data from the included studies were extracted independently and in duplicate using a standardized data collection form [24]. Two reviewers independently assessed studies using the Cochrane risk of bias tool [47] and the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system [48].

Data Synthesis and Analysis

Meta-analyses for primary and secondary outcomes were conducted using random effects models with standardized mean difference (SMD) and mean difference (MD) where appropriate in Review Manager (RevMan; version 5.4, The Cochrane Collaboration, 2020) [24]. According to the Cochrane handbook, when necessary, we converted scales to correct for the difference in direction [49], and median and IQR were converted following the methods of Wan et al [50]. When possible, we performed sensitivity analyses by removing studies with an overall rating of a high risk of bias for each outcome. This deviates from the initial protocol, wherein we planned to remove only studies with a high risk of bias in ≥3 domains. Where appropriate, we completed subgroup analyses for our secondary questions. Where possible, we completed analyses for interventions that used smart technology alone compared with face-to-face alone.

Results

Overview

We identified 12,245 records from our search; reviewers screened 9434 titles and abstracts after duplicate removal, and 19 RCTs were eligible for inclusion (Figure 1). The reasons for full-text exclusion are provided in Multimedia Appendix 1. Reviewers attempted to contact 7 corresponding authors for missing information, and 1 author responded. All the studies were combined in a narrative synthesis, evaluating a total of 3455 participants. A total of 18 RCTs with 3405 participants randomized to either a smart technology or face-to-face intervention were included in the quantitative analysis.

Characteristics of the studies are shown in Tables 1-2. Of the 3455 participants, 1874 (54.24%) were female, and the mean age ranged from 60 to 72 years. Studies were conducted at outpatient or community practices in the following countries: United States (6/19, 32%); Belgium (2/19, 11%); Netherlands (2/19, 11%); United Kingdom (2/19, 11%); and one each in Australia, Chile, Denmark, Finland, Hong Kong, New Zealand, and Spain. Participants were community-dwelling older adults, with 14 studies focused on specific clinical populations, including people with chronic obstructive pulmonary disease (COPD) [27-31], cardiovascular disease [32-35], diabetes [36,37], knee arthritis [38,39], obesity [40], and cognitive impairment plus physical frailty [41].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram of evidence search and selection.

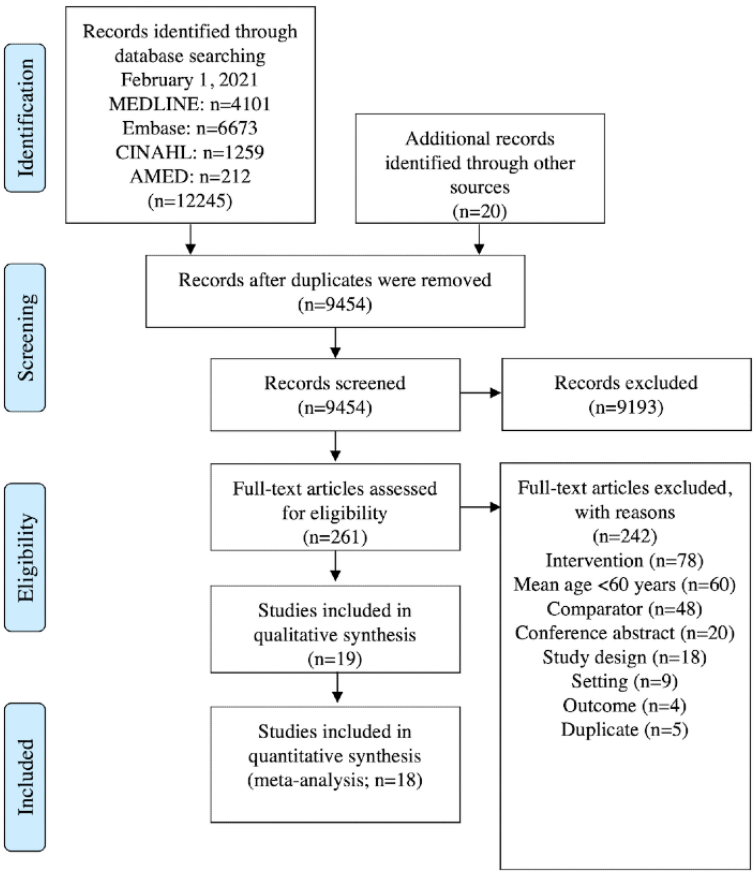


Table 1. Characteristics of included studies with author names beginning with A to J.

Study and country	Population, sample size (n)	Intervention[a]	Usual care	Outcomes	Key findings
Alonso-Domínguez et al [36], 2019, Spain	T2DM[b], median age IG[c]=60.8 (IQR 7.8) years; CG[d]=60.4 (8.4) years, 45.6% female, n=204	[Smartphone app] Daily use of mobile app for 3 months + heart-healthy 4-km walks once per week for 5 weeks	10-minute standardized counseling session on PA[e] and healthy diet; information leaflet provided	PA: steps per day measured by pedometer; weekly PA measured by IPAQ[f]-Spanish version	Significant increase in daily steps per day (P<.05 at 3 and 12 months), aerobic steps (P<.05 at 3 months), distance walked (P<.05 at 3 months), and total PA (P<.05 at 3 months) for intervention group
Armit et al [43], 2005, Australia	Community-dwelling older adults (55-70 years), mean age 64 (SD 4.6) years, 64.9% female, n=37 recruited, 28 with 1 assessment	[Pedometer] Pedometer use for 12 weeks; counseling session; information booklet; diary for goal setting and self-monitoring; 3 follow-up phone calls to reevaluate, reinforce, and discuss adherence issues	Counseling session; information booklet; diary for goal setting and self-monitoring; 3 follow-up phone calls to reevaluate, reinforce, and discuss adherence issues	PA: self-report survey[g]	No significant difference in PA between groups at either time point
Audsley et al [44], 2020, United Kingdom	Community-dwelling older adults from Falls Management program, mean age IG=76.9 (SD 7.0) years; CG=73.8 (SD 6.4) years, 73.3% female, n=50	[Pedometer] 6 sessions of 60 to 90 minutes of motivational interviewing and behavior change techniques over 6 months; pedometer use worksheets with PA diaries + usual care	24-week group exercise program (strength, balance, cardiovascular, and flexibility exercises) + 30-minute 2 times per week home exercise program; how to get up from a fall and strategies to reduce inactivity	PA: weekly MVPA[h] measured by Phone-FITT[i] questionnaire	No significant differences in MVPA between groups
Barnason et al [32], 2009, United States	Postoperative CABS[j] (aged >65 years), mean age 71.2 (SD 4.9) years, 17% female, n=280	[Telerehabilitation] Telehealth symptom management for 6 weeks on strategies to address common symptoms after CABS and improve outcomes (eg, PA and functioning) + usual care	Usual care not defined	PA: modified 7-day Activity Interview; activity counts by accelerometer; Physical Activity and Exercise Diary; quality of life: MOS SF-36[k]	Higher overall energy expenditure in usual care group, except at 3 weeks where intervention group had higher expenditure; no other significant differences
Christiansen et al [38], 2020, United States	Unilateral TKR[l] (aged >45 years), mean age 67 (SD 7.0) years, 53.4% female, n=43	[Wearable] 6-month standard outpatient PT[m] with a Fitbit Zip, weekly steps per day goal from a PT, 6 monthly phone calls from a research assistant	Standard outpatient PT, including printed home exercise program; 6 monthly phone calls after discharge discussing overall health	PA: steps per day and MVPA per week measured by Actigraph	Increased steps and minutes per week in MVPA for intervention group at both follow-ups
de Blok et al [27], 2006, Netherlands	COPD[n] (aged 40-85 years), mean age IG=65 (SD 10.4) years, CG=62.5 (SD 12.3) years, 75% female, n=21	[Pedometer] 4 sessions of 30-minute exercise counseling (steps per day goal setting), pedometer feedback for 10 weeks + usual care	Conventional pulmonary rehabilitation of exercise training, dietary intervention, and psychoeducational modules over 9 weeks	PA: pedometer steps per day[g]; function: chair stand, arm curl, 8-foot up-and-go, 2-minute step test; other: Beck Depression Inventory	Significantly improved 8-foot up-and-go test and 2-minute step test for intervention group; no significant differences between groups in steps per day
Frederix et al [33], 2015, Belgium	Cardiac conditions (CAD[o] and CHF[p]), mean age 61 (SD 8.5) years, 17% female, n=140	[Multicomponent] 24-week PA telerehabilitation; exercise training protocols + center-based cardiac rehabilitation for 12 weeks (usual care)	Center-based cardiac rehabilitation for 12 weeks (45 sessions—2 exercise sessions per week, dietary consultation, and psychologist session)	PA: CPET[q] VO_2[r] peak[g], accelerometer PA, self-report IPAQ converted to MET[s]-minutes per week of MVPA; quality of life: HeartQoL	Significant improvement in mean VO_2 peak (P<.001), steps per day, MET-minutes per week of MVPA (P=.01), and health-related QoL[t] (P=.01) for intervention group

Study and country	Population, sample size (n)	Intervention[a]	Usual care	Outcomes	Key findings
Hansen et al [28], 2000, Denmark	COPD (no participation in pulmonary rehabilitation in the preceding 6 months), mean age 68.3 (SD 9.0) years, 55% female, n=134	[Telerehabilitation] 10-week telerehabilitation program including group-based supervised, standardized 35-minute exercise program via videoconference, followed by 20 minutes of patient education 3 times per week	10-week conventional pulmonary rehabilitation program including group-based supervised, standardized 60-minute exercise program 2 times per week in conjunction with 1 time per week education session lasting 60-90 minutes	PA: accelerometer PA; function: 6MWT[g,u], 30-second sit-to-stand; quality of life: HADS[v], EQ-5D[w]; COPD: COPD assessment test and Clinical COPD Questionnaire	No significant differences in 6MWT; however, intervention group maintained improvement at 22 weeks; significantly improved HADS and depression scores (P<.05) and COPD Assessment Test scores (P=.04) in intervention group

[a]The authors grouped interventions by the type of smart technology; groupings are indicated in square brackets at the beginning of each intervention description.

[b]T2DM: type 2 diabetes mellitus.

[c]IG: intervention group.

[d]CG: control group.

[e]PA: physical activity.

[f]IPAQ: International Physical Activity Questionnaire.

[g]Primary outcomes of the individual studies.

[h]MVPA: moderate to vigorous physical activity.

[i]FITT: Frequency, intensity, type, and time

[j]CABS: coronary artery bypass surgery.

[k]MOS SF-36: Medical Outcomes Study 36-item Short Form survey.

[l]TKR: total knee replacement.

[m]PT: physiotherapy.

[n]COPD: chronic obstructive pulmonary disease.

[o]CAD: coronary artery disease.

[p]CHF: congestive heart failure

[q]CPET: cardiopulmonary exercise testing.

[r]VO_2: maximal oxygen uptake.

[s]MET: metabolic equivalent.

[t]QoL: quality of life.

[u]6MWT: 6-minute walk test.

[v]HADS: Hospital Anxiety and Depression Scale.

[w]EQ-5D: EuroQoL 5 Dimensions.

Table 2. Characteristics of included studies with author names beginning with K to Z.

Study and country	Population, sample size (n)	Intervention[a]	Usual care	Outcomes	Key findings
Kawagoshi et al [29], 2015, Finland	COPD, mean age IG=74 (SD 8) years; CG=75 (SD 9) years, 11% female, n=39	[Pedometer] Pedometer + home-based pulmonary rehabilitation including breathing retraining, exercise training, respiratory training, and monthly 45-minute education sessions	Home-based pulmonary rehabilitation including breathing retraining, exercise training, respiratory muscle training, and monthly 45-minute education sessions	PA: total PA as measured by accelerometer; function: quadriceps muscle force; 6MWT; Pulmonary: Chronic Respiratory Disease Questionnaire	Significant increase in walking time (*P*=.04), dyspnea, and quadriceps force in intervention group; significant improvements in pulmonary function tests, 6-minute walk distance, and Chronic Respiratory Disease Questionnaire in both groups
King et al [42], 2020, United States	Community-dwelling adults aged 50 years, mean age 62.3 (SD 8.4) years, 78.8% female, n=245	[Telerehabilitation] 1:1 initial counseling session, pedometer use; additional (up to 28) 10 to 15 minutes counseling sessions by a virtual advisor over 12 months	1:1 initial counseling session, pedometer use; additional (up to 28) 10 to 15 minutes counseling sessions conducted by a human advisor over 12 months	PA: walking minutes per week[b] total PA, MVPA, daily PA measured by accelerometer, self-report weekly sedentary behavior; quality of life: Vitality Plus scale	Significantly increased steps per day (*P*=.02) in intervention group; significant decreases in both groups for reported sedentary time
Kwan et al [41], 2020, Hong Kong	Mild cognitive impairment and physical frailty, mean age 71.0 (SD 9.0) years, 85% female, n=33	[Smartphone app] Smartphone apps for individualized goals, to log PA data, performance reviews, and e-reminders; communication app for e-coaching, personalization of goal settings, and messages of praise + control intervention; all interventions were for 12 weeks	Conventional behavior change techniques via PA counseling, telephone follow-up, health education, and exercise training. All interventions were for 12 weeks	PA: steps per day measured by accelerometer; MVPA and non-MVPA per week measured by accelerometer; other: Fried frailty index; MoCA[c]	Significant increase in MVPA (*P*=.04), walking time (*P*=.03), steps per day (*P*=.02), brisk walking (*P*=.009), and peak cadence (*P*=.003) for intervention group; adherence to face-to-face sessions was 100% for both groups; smartphone compliance was 54.1 (SD 1.2) days per participant (range 0-56 days)
Maddison et al [34], 2015, New Zealand	Ischemic heart disease, mean age 60.2 (SD 9.3) years, 19% female, n=171	[Multicomponent] Automated text messages for 24 weeks encouraging 30 minutes per day MVPA 5 days per week + regular exercise prescription, behavior change strategies, website access with model vignettes, self-monitoring, information	Encouraged to participate in cardiac rehabilitation typically including education sessions, psychological support, and PA encouragement and offer to join a supervised exercise club	PA: self-reported PA measured by IPAQ; other: PVO_2[d] assessed during CPET[b]; self-efficacy and motivation to exercise; SF-36 and EQ-5D	No significant difference in PVO_2 between groups; significant improvements in self-report PA (*P*=.05), walking (*P*=.02), self-efficacy (*P*=.04), and health-related QoL (*P*=.03) in intervention group
Mendoza et al [30], 2020, Chile	Stable COPD, mean age 68.7 (SD 8.5) years, 39.2% female, n=102	[Pedometer] Pedometer use and steps per day goals for 3 months, 3 monthly follow-up sessions with a physician and physiotherapist to increase step count	3 monthly counseling sessions with physician and PT to increase PA, advised to walk minimum 30 minutes per day	PA: steps per day[b] measured by pedometer; other: health status and exercise capacity	Significant increase in PA (*P*<.001) and exercise capacity (*P*=.03) in intervention group
Mouton and Cloes [45], 2015, Belgium	Community-dwelling adults (aged 50 years), mean age IG1=61.2 (SD 6.3) years; IG2=69.8 (SD 7.4) years; IG3=63.2 (SD 5.7) years; CG=66.1 (SD 6.8) years, 60% female, n=149	[Website] Three groups, all 3-month duration: (1) web-based intervention (PA promotion + monthly PA feedback); (2) center-based intervention—12 weekly sessions of group exercise; and (3) mixed intervention (web- and center-based intervention)	No intervention received	PA: self-report using IPAQ-S[e]; other: stages of change; awareness of PA; and participant acceptance of intervention	Mixed intervention increased PA level (*P*=.04); center-based intervention (*P*<.001) and mixed intervention (*P*=.01) increased PA stages of change; web-based intervention (*P*=.02) and mixed intervention (*P*<.001) increased PA awareness

Study and country	Population, sample size (n)	Intervention[a]	Usual care	Outcomes	Key findings
Roberts et al [35], 2019, United States	Community-dwelling adults (aged 60 years) with moderate to high risk of CVD[f] events, mean age 72 (SD 7.4) years, 60% female, n=40	[Wearable] Activity tracker + strategies to increase PA for 20 weeks; usual care (8-week center-based exercise intervention); goal to achieve 150 minutes of MVPA per week for 12 weeks; encouragement of nonexercise PA	Usual care including 2 for per week center-based exercise intervention for 8 weeks; instruction to achieve 150 minutes of MVPA per week for remaining 12 weeks; behavioral counseling to encourage nonexercise PA	PA: daily activity measured by accelerometer; function: 6MWT; 4-meter gait speed; grip strength; SPPB[g]	Significant increase in steps per day for intervention group
Tabak et al [31], 2014, Netherlands	Stable COPD, mean age IG=65.2 (SD 9.0) years; CG=67.9 (SD 5.7) years, 37% female, n=34	[Multicomponent] 4-week daily use of mobile activity coach for feedback, motivation, and target PA levels + usual care (medication and PT—weekly group training sessions)	Could consist of medication and weekly group training PT sessions	PA: steps per day[b] measured by pedometer; COPD: Clinical COPD Questionnaire (health status); other: compliance	No significant differences in steps per day; nonsignificant improvement in health status in intervention group; 86% adhered to the activity coach
Talbot et al [39], 2003, United States	Symptomatic knee OA[h], aged ≥60 years, mean age IG=69.6 (SD 6.7) years; CG=70.8 (SD 4.7) years, 76.5% female, n=34	[Pedometer] Pedometer use + daily step goals; education booklet on exercise and managing pain; usual care (12 sessions of 1-hour arthritis self-management education)	12 sessions of 1-hour arthritis self-management education (including a session on exercise)	PA: steps per day[b] by pedometer, PA over time (accelerometer); function: leg muscle strength; 100-foot timed walk-turn-walk; timed stair climb; timed chair rise	23% steps per day increase for intervention group vs 15% decrease in control group; improved usual pace gait speed (*P*=.04) and isometric leg strength (21%—compared with 3.5% loss in control group)
Weinstock et al [37], 2011, United States	Diabetes mellitus, mean age 70.9 (SD 6.8) years, 63% female, n=1650	[Multicomponent] Educational videoconferencing for 4-6 weeks to review blood glucose and blood pressure measurements; pedometer use with goals set for 2 years	Usual care from PCP[i]; PA encouraged by pedometer use with goals set between participant and PCP for 2 years	PA: diabetes self-care activities for assessment of PA; other: feasibility; acceptability; and CARE Depression Instrument	Significantly slower rate of decline in PA (*P*=.01) and lower rate of PI[j] (0.04); significantly higher PA levels (*P*<.001) in intervention group
Yates et al [40], 2009, United Kingdom	Overweight or obese (BMI ≥25), mean age 65 (SD 8) years, 34% female, n=87	[Pedometer] Pedometer use + 180-minute education session on causes and complications of impaired glucose tolerance + exercise information	Two groups: (1) Same education session as intervention but no pedometer and (2) usual care—information pamphlet	PA: steps per day measured by pedometer, self-reported walking by IPAQ, and total MVPA	Compared with usual care group 2, significant increases in steps per day, self-reported walking, and total MVPA at 3, 6, and 12 months in intervention group (all *P*<.05)

[a]The authors grouped interventions by the type of smart technology; groupings are indicated in square brackets at the beginning of each intervention description.

[b]Primary outcomes of the individual studies.

[c]MoCA: Montreal Cognitive Assessment.

[d]PVO_2: peak oxygen uptake.

[e]IPAQ-S: International Physical Activity Questionnaire-Short.

[f]CVD: cardiovascular disease.

[g]SPPB: Short Performance Physical Battery.

[h]OA: osteoarthritis.

[i]PCP: primary care provider.

[j]PI: physical impairment.

Interventions

Detailed descriptions of the smart technology interventions in each study are provided in Multimedia Appendix 1. A total of 16 studies included a single smart technology component, including smartphone apps [36,41], wearable activity trackers (eg, Fitbit) [35,38], telerehabilitation (eg, video conferencing, virtual advisor, or health buddy device) [28,32,42], and pedometers (ie, only provides step counts) [27,29,30,39,40,43,44], whereas the remaining 3 had multiple components, including video conference and pedometer [37]; website and SMS text messaging [34]; and website plus

pedometer, SMS text messaging, and email [33]. A total of 15 studies included smart technology and a face-to-face component in the intervention group [27,29-36,38-41,43,44]. Furthermore, 4 studies evaluated smart technology alone versus face-to-face alone [28,37,42,45]. The length of interventions ranged from 10 weeks [27] to 1 year [29], and follow-ups ranged from 6 weeks [32] to 2 years [37].

Risk of Bias

Overall, the risk of bias was a concern for studies included in this review. On the basis of the outcome with the highest risk of bias (ie, if we assessed risk of bias for 3 outcomes in a study and 1 of those was rated a high risk of bias and the other 2 were some concerns, we rated the study as high risk of bias), 14 studies were judged to be high risk [27,29,31,32,34,36,37,39-45], 4 studies had some concerns [28,30,35,38], and 1 study had a low risk of bias [33]. The areas of greatest concerns were missing outcome data (10/19, 53% high risk) and risk of bias related to measurement of the outcome (eg, awareness of intervention and influence of knowledge of intervention on the assessment; 9/19, 47% high risk). A summary figure is available in Multimedia Appendix 1.

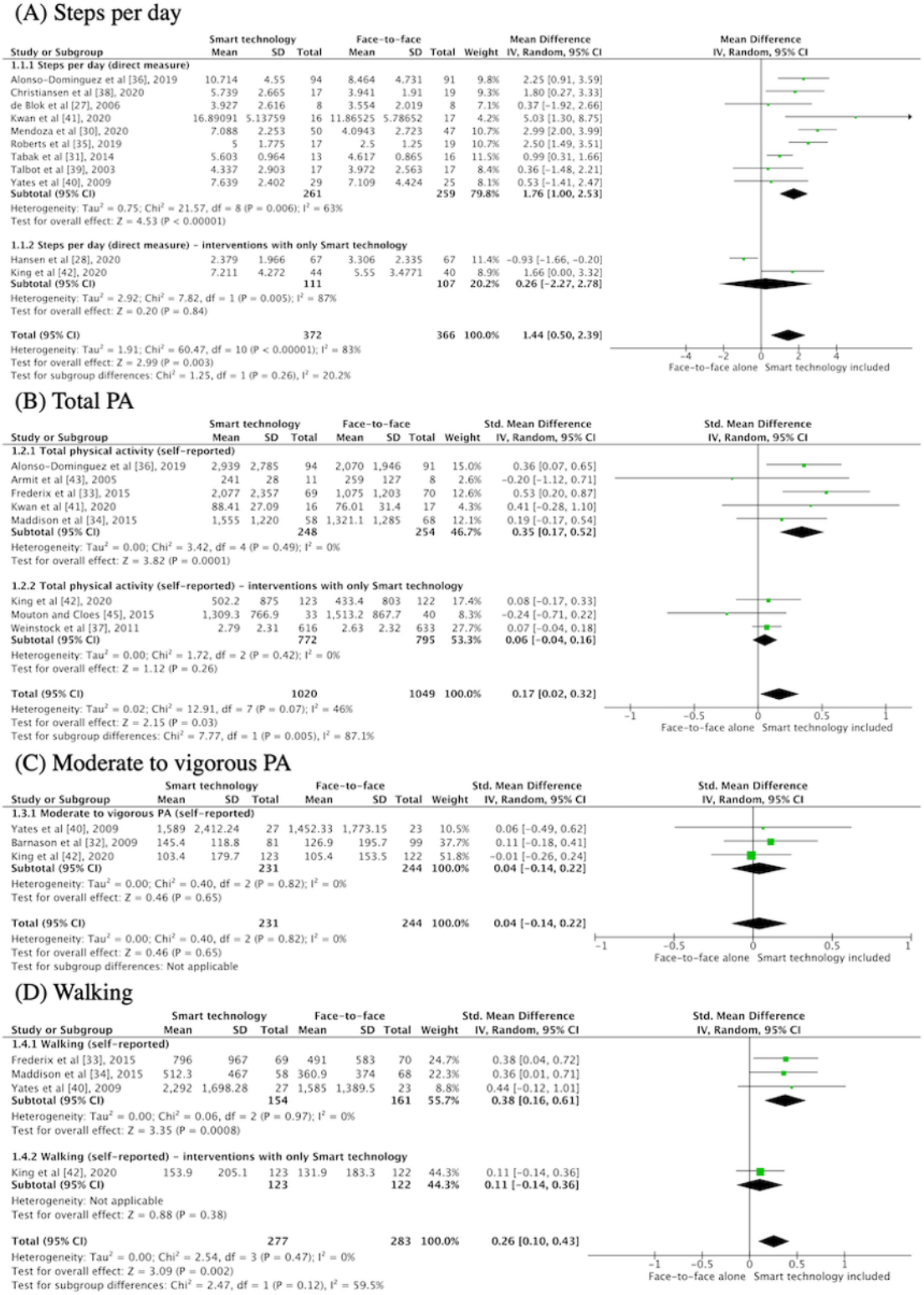
Physical Activity

Overview

All included studies assessed the effect of smart technology interventions on PA [27-45]. The types of PA evaluated included steps per day [27,28,30,33,35,36,38-40,42], total PA [28,33,34,36,37,41-43,45], moderate to vigorous PA [32,38,40-42,44], and walking [29,33,34,40-42], assessed either directly (eg, with pedometer or activity tracker) or indirectly (eg, self-report measures such as the International Physical Activity Questionnaire [33,34,36,40,45], Diabetes Self-Care Activities for assessment of PA [37], Community Health Activities Model Program for Seniors PA [42], Physical Activity Scale for the Elderly [41], activity diary [32], and Active Australia Survey [43]). Studies were grouped by type of PA, and 4 meta-analyses were performed for daily step counts, total PA, moderate to vigorous PA, and walking (Figure 2).

Compared with face-to-face interventions, interventions that included smart technology improved step count, with the meta-analysis of 11 studies and 738 participants demonstrating a MD of 1440 steps (95% CI 500-2390; Figure 2) [27,28,30,31,35,36,38-42]. Smart technology also improved total PA scores (8 studies, n=2069; SMD 0.17, 95% CI 0.17-0.52) and walking (4 studies, n=560; SMD 0.26, 95% CI 0.10-0.43) compared with face-to-face interventions. The meta-analysis of 3 studies (n=475) for moderate to vigorous PA was not statistically significant (SMD 0.04, 95% CI −0.14 to 0.22). We performed sensitivity analyses for all outcomes except moderate to vigorous PA. On the basis of the results of the risk of bias for the sensitivity analyses, we removed 5 studies with high risk of bias in the steps per day; however, the remaining 6 studies still favored smart technology with a MD of 2.03 (95% CI 0.35-3.71). Only 2 studies remained with 1388 participants for total PA; the difference was no longer significant with a SMD of 0.27 (95% CI −0.18 to 0.72). Finally, our sensitivity analysis for walking containing 2 studies (n=384) was not significant with a SMD of 0.22 (95% CI −0.04 to 0.48).

Figure 2. Meta-analysis of effect of smart technology versus face-to-face alone on physical activity (PA).



Subgroup Analyses

Smart Technology Components

We completed meta-analyses for subgroups according to the type of smart technology that was used for steps per day and total PA scores (Multimedia Appendix 1, section 5). Studies assessing steps per day used pedometers [27,30,39,40], smartphone apps [36,41], telerehabilitation [28,42], and wearable activity trackers in their interventions [35,38]. When examining the effects of each intervention component separately, only smartphone apps (MD 3.07, 95% CI 0.5-5.55) and wearable activity trackers (MD 2.29, 95% CI 1.44-3.13) showed significant improvements in favor of smart technology interventions. Smart technology interventions for studies that assessed total PA included pedometers [43], telerehabilitation [42], websites [45], smartphone apps [36,41], and interventions including multiple smart technology components [33,34,37]. In the subgroup meta-analyses, smartphone apps showed

significant improvements in total PA scores compared with face-to-face alone (SMD 0.37, 95% CI 0.10-0.63), and multicomponent interventions favored smart technology, but the effect was not statistically significant (SMD 0.23, 95% CI −0.05 to 0.51).

Smart Technology Alone

There was a large variability in the components making up smart technology interventions in the included studies. Only 4 RCTs (n=1701) evaluated the effect of an entirely smart technology alone intervention (ie, did not include any in-person consultations) versus face-to-face alone [28,37,42,45]. Subgroup analyses were performed for both steps per day and total PA scores. For both outcomes, the pooled results were not significant and did not appear to favor either intervention (Figure 2).

The evidence for the effect of smart technology interventions on PA was judged to be very low based on the GRADE criteria (Figure 3). Therefore, our confidence in the effect estimate is limited, and the true effect is likely to be different from the estimated effect.

Figure 3. Grading of Recommendations Assessment, Development, and Evaluation (GRADE) summary of findings table.

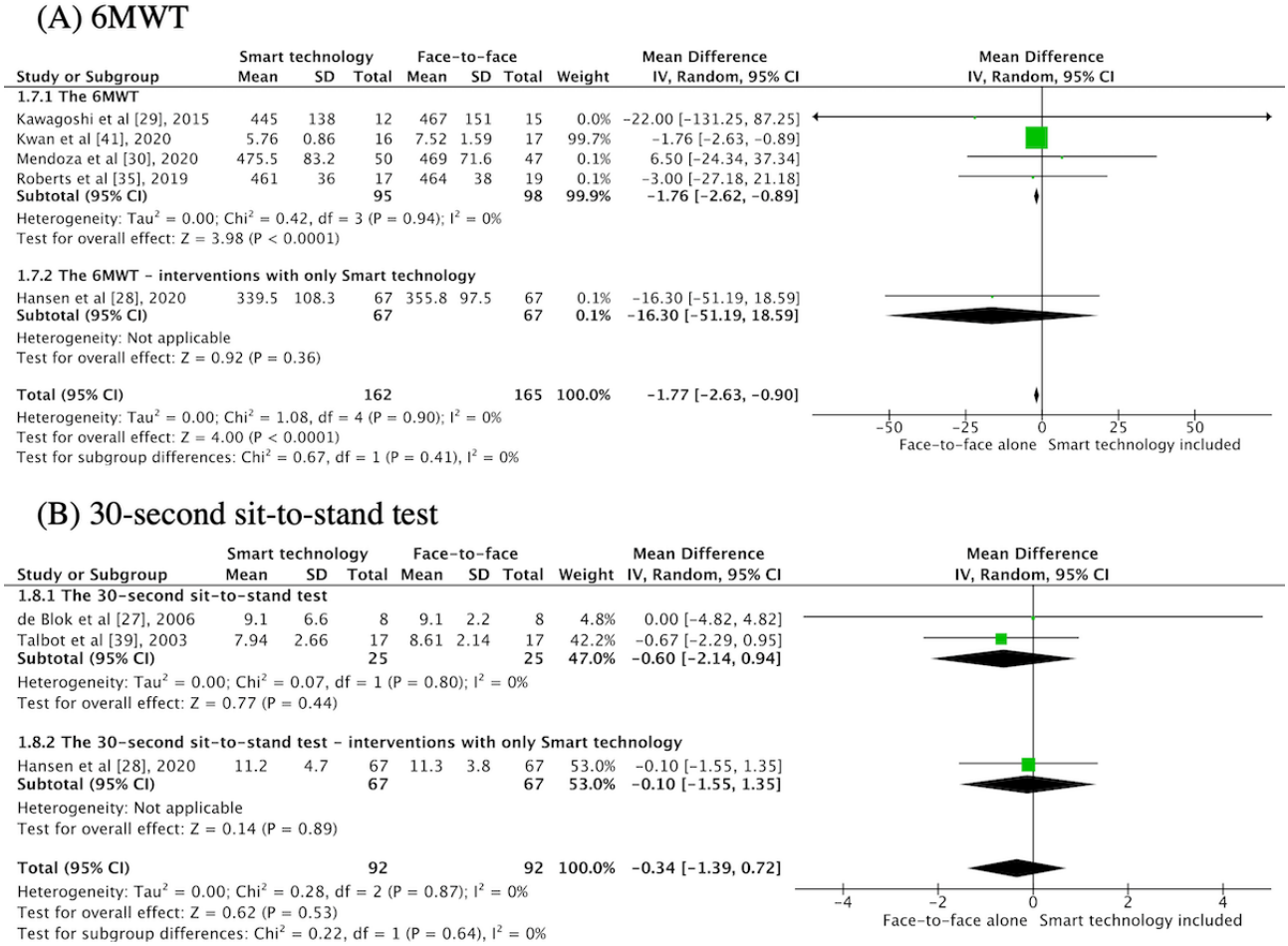
Outcomes	Number of participants	Number of studies	Certainty of evidence (GRADE)	Comments
Physical Activity				
Steps per day	877	12/18 (67%)	⨁◯◯◯	A total of 11 studies showed improvement in steps per day, favoring smart technology interventions. Certainty of evidence is very low owing to risk of bias, inconsistency, indirectness, and imprecision.
Self-reported total physical activity	2036	9/18 (50%)	◯⨁◯◯	A total of 6 studies showed increased levels of total physical activity, favoring smart technology interventions. Certainty of evidence is low owing to high risk of bias, indirectness, and imprecision.
Self-reported walking time	620	6/18 (33%)	⨁◯◯◯	All 6 studies showed increased walking time, favoring the smart technology interventions. Certainty of evidence is very low owing to risk of bias, indirectness, and imprecision.
Moderate to vigorous physical activity	585	6/18 (33%)	⨁◯◯◯	A total of 4 studies showed increased time in moderate to vigorous physical activity, favoring smart technology interventions. Certainty of evidence is very low owing to risk of bias, indirectness, and imprecision.
Physical Function				
Physical function – 6MWT[a]	347	6/18 (33%)	◯⨁◯◯	A total of 2 studies showed improvements in 6MWT, favoring smart technology interventions. Certainty of evidence is low owing to high risk of bias, indirectness, and imprecision.
Physical function – 30-second sit-to-stand test	184	3/18 (17%)	◯◯⨁◯	There was no difference in the sit-to-stand outcomes between smart technology interventions and controls. Certainty of evidence is moderate owing to risk of bias and imprecision.
Adherence	434	9/18 (50%)	⨁◯◯◯	A total of 8 studies demonstrated good adherence to the smart technology interventions, with 1 study unable to determine owing to missing data. Certainty of evidence very low due to risk of bias, indirectness, and imprecision.
Secondary Outcomes				
Quality of life	442	6/18 (33%)	⨁◯◯◯	A total of 2 studies showed improvements in quality of life, favoring smart technology. Certainty of evidence very low due to risk of bias, indirectness, and imprecision.
Depression	150	2/18 (1%)	⨁◯◯◯	Both studies showed improvements in depression scales favouring smart-technology interventions. Certainty of evidence very low due to risk of bias, indirectness, and imprecision.

Physical Function

Overview

Of 19 studies, 10 (53%) evaluated some aspects of physical function. The most common performance-based measures were the 6-minute walk test (6-MWT) [28-30,35,41] and the 30-second sit-to-stand test [27,28,39]. The remaining 13 measures were only included in 1 or 2 studies (eg, 4-meter gait speed [34], short physical performance battery [34], timed stair climbing [35], peak VO_2 [14,39], hand grip strength [31,34], maximal inspiratory and expiratory force [29], and quadriceps force) [29,35] and therefore were not pooled (Figure 4). In total, 5 RCTs were pooled in a random effects meta-analysis for 6-MWT including 327 participants, with a MD of −1.77 m (95% CI −2.63 to −0.90) [28,29,32,34,38]. Furthermore, 3 studies (n=184) that used the 30-second sit-to-stand test found no significant difference between interventions (SMD −0.37, 95% CI −1.66 to 0.92). The sensitivity analysis for the 6-MWT test included 3 studies (n=267). After removing those with a high risk of bias, the difference was not statistically significant (MD −3.26, 95% CI −19.97 to 13.45; Multimedia Appendix 1).

Figure 4. Meta-analysis of smart technology versus face-to-face alone for physical function. 6MWT- 6-minute walk test.



Notes: green squares represent individual study effect sizes; black diamonds represent pooled effect sizes of subgroups and total effect size.

Subgroup Analyses

Smart Technology Intervention Components

We conducted additional meta-analyses on 5 studies that included the 6-MWT as an outcome by type of smart technology intervention component: pedometers [29,30], smartphone apps [41], telerehabilitation [28], and wearable activity trackers [35]. We examined the effects of pedometer interventions on the 6-MWT on their own. There was a nonsignificant MD (4.40, 95% CI −25.28 to 34.07). Smart technology components included in the 3 studies with the 30-second sit-to-stand test were pedometers [27,39] and telerehabilitation [28]. The findings were nonsignificant for pedometers compared with face-to-face alone with an MD of −0.60 (95% CI −2.14 to 0.94). Complete results for these meta-analyses are available in section 6 of Multimedia Appendix 1.

Smart Technology Alone

Only 1 study (n=134) examined a smart technology intervention alone using telerehabilitation (video conferencing) [28]. For both the 6-MWT and the 30-second sit-to-stand test, results in this study favored the face-to-face intervention.

The evidence for the effect of smart technology on the 6-MWT and the 30-second sit-to-stand test was judged to be very low based on the GRADE criteria. Therefore, our confidence in the effect estimates is limited, and the true effect may be substantially different from the estimated effect.

Adherence Rates

A total of 10 studies did not report adherence rates [30,32-34,37,38,40,43,45]. Two studies reported overall adherence to the intervention [28,42], and 7 studies reported adherence to the components of the intervention [27,29,31,35,39,41,44]. Overall, across the available studies, adherence rates to smart technology interventions ranged from 64.8% [42] to 95.5% [27]. Adherence to specific smart technology interventions was reported as follows: pedometers ranged from 76% [39] to 95.5% [27], and telerehabilitation ranged from 64.8% [32] to 85% [28]. The adherence to face-to-face interventions ranged from 64% [28] to 100% [41].

Secondary Outcomes

Three studies measured depression using different scales, including the Hospital Anxiety and Depression Scale [28], Beck Depression Inventory [27], and Care Depression Instrument [37]. We pooled data from 2 RCTs for a random effects meta-analysis, which showed no significant difference between interventions that included smart technology and face-to-face alone (SMD 0.09, 95% CI −0.23 to 0.41) [27,28]. Furthermore, 6 studies measured HRQoL using the Clinical COPD Questionnaire [28,31], St George's Respiratory Questionnaire [27,30], HeartQoL global score [33], and Chronic Respiratory Disease Questionnaire [29]. A meta-analysis of 6 studies demonstrated a significant SMD (0.31, 95% CI 0.11-0.51) in favor of smart technology interventions. For HRQoL, we completed a subgroup analysis by smart technology component; multicomponent interventions (n=2; SMD 0.35, 95% CI 0.05-0.66) showed significant improvements compared with face-to-face alone, but pedometers did not (n=3; SMD 0.04, 95% CI −0.49 to 0.57). After performing a sensitivity analysis excluding studies with a high risk of bias, our results did not change the significance or direction of results. Only 1 study examined anxiety using the Hospital Anxiety and Depression Scale and demonstrated an improvement in favor of smart technology at the first follow-up [28]. Meta-analyses for secondary outcomes are available in Multimedia Appendix 1.

Discussion

Principal Findings

PA is integral to reducing age-related illness and disability. This review of 19 studies involving 3455 patients found that interventions that include smart technology may improve steps per day, total PA, walking, and HRQoL in older adults. Although the overall quality and certainty of the evidence were judged to be very low and more precise estimates will need to be obtained, our results may have important implications for research and practice on PA promotion, especially in the context of the COVID-19 pandemic and public health restrictions.

Among the included studies, we found that most (n=15) smart technology interventions used multiple components with intervention groups also receiving a face-to-face component. One of the challenges of multicomponent interventions is that it limits our understanding of the effectiveness of smart technology alone. For example, intervention and control groups in several studies received usual care, with the only difference between groups being the addition of a pedometer or wearable activity tracker [29,38-40]. Of the 4 studies examining smart technology alone, 2 examined the effect on steps per day; our subgroup analysis found a nonsignificant difference of 260 steps per day favoring smart technology [28,42]. Notably, the results from these 2 studies were contradictory, with Hansen et al [28] reporting that face-to-face interventions were more effective. However, they evaluated patients with severe COPD and reported a decline in PA for both groups over the course of the study. This may be caused, in part, by the progressive nature of COPD or by the short length of the intervention at only 10 weeks [28]. In line with our meta-analysis for steps per day, the 3 RCTs examining the effect of smart technology alone on total PA showed a small, nonsignificant effect favoring smart technology [37,42,45]. It will be important for future work to determine whether smart technology on its own is more, less, or as effective as face-to-face interventions. This is especially important during the COVID-19 pandemic, as in-person contacts should be limited to minimize the risk of transmission. Conversely, if smart technology interventions alone are less effective, it will be important to revert to including some type of face-to-face component as soon as it is safe and feasible.

Patient-important outcomes were not well represented in our included studies, which highlights an important gap in the design and reporting of smart technology interventions. Among 19 studies, only 7 (37%) unique studies reported on patient-important outcomes: 6 reported on HRQoL [27-31,33], 3 on depression [27,28,37], and 1 on anxiety [28]. Evidence has demonstrated associations between PA levels and mental health and quality of life in older adults [51]. Therefore, it is vital that research includes these measures to further our understanding of the magnitude and direction of these relationships. This is timely as we have seen the prevalence of depression and anxiety increasing over the last 2 years with the pandemic and public health restrictions [52]. Older adults have been shown to be at an increased risk for anxiety [52]. In addition to the lack of patient-important outcomes, we found that data on intervention adherence were largely missing or inconsistently reported. This leads to several challenges, including understanding the acceptability of these interventions for older adults and interpreting results of the primary studies. For example, if adherence to the intervention was poor, it would be difficult to appreciate if the differences (or lack thereof) between groups were because of intervention failure or implementation failure. The issue of acceptability is also crucial, particularly considering that the current literature on the usability of technology in this population is limited.

Our results suggest that the type of smart technology used for PA interventions may influence the effectiveness of the intervention. We found that studies that used smartphone apps led to significant improvements in steps per day and total PA scores [53]. Other systematic reviews examining the effectiveness of mobile phone interventions in adults aged >18 years and >50 years and older adults (>65 years) have shown mixed evidence for improving PA [53-55]. Potential reasons for these discordant results may be the diversity of interventions within control groups, small sample sizes, and moderate to high levels of heterogeneity across studies [53,54]. We also found

that wearable activity trackers significantly improved the number of steps per day. Researchers should consider the type of smart technology in conjunction with the PA goals and the population to achieve the best outcome. Further research is warranted to determine the optimal types of smart technology interventions for older adults.

The existing literature has several limitations that warrant future research. Our findings are based on very low quality of evidence, as per the GRADE criteria. All included studies had some concerns or a high risk of bias, most commonly because of missingness of outcomes, lack of intention-to-treat analyses, inadequate allocation concealment, and lack of prespecified statistical analyses or protocols. The variability in populations, smart technology interventions, control group interventions, and outcomes among studies are major contributing factors to the large degree of statistical heterogeneity. Furthermore, most studies used multiple components, making it difficult to assess which parts may have contributed to differences between groups or, conversely, if additive components may have diluted a potentially effective intervention. Importantly, 79% (15/19) of the studies included an element of face-to-face interaction in the intervention group, making it difficult to evaluate the effectiveness of smart technology alone. Due to the limited number of studies and small sample sizes, there is a need to group studies and smart technologies broadly to have sufficient sample sizes to conduct meta-analyses. Given that different smart technologies can be used for different intervention components, it may be difficult to apply our findings in practice when designing interventions. Although we conducted subgroup analyses where there were sufficient data, there is a need for additional research comparing different types of smart technologies for supporting specific PA interventions in older adults [18,19,23]. Finally, our results may have been influenced by the high risk of performance bias caused by the impracticability of blinding therapists and participants owing to the nature of the interventions. Future research should focus on minimizing the risk of bias, evaluating individual smart technology components and smart technology alone, and including standardized control groups. Improved reporting of control group interventions may also assist with interpretation of results.

This review also has some limitations. We did not include gray literature owing to a lack of central sources to identify and retrieve these citations [56]. In addition, we excluded studies published in languages other than English because of the feasibility of the review. Therefore, our cohort of studies may not represent the entirety of the literature. This review has several important strengths. To the best of our knowledge, this is the first study to attempt to compare smart technology interventions specifically with face-to-face interventions, which is critical for determining their effectiveness compared with traditional modes of delivery. In addition, we published our peer-reviewed protocol, and we developed and conducted our search in collaboration with a health research librarian.

Conclusions

In the context of substantial heterogeneity and very low quality of evidence, our results suggest that PA interventions that include smart technology components may significantly improve steps per day and total PA in community-dwelling older adults. Subgroup analyses showed that smartphone apps and wearable activity trackers seem to be the most effective smart technology, which may be helpful for health care practitioners when determining appropriate methods of remote PA promotion. When comparing smart technology alone with face-to-face alone, there were few studies with discordant results and no significant differences between groups. The results should be interpreted with caution given the challenges with the existing literature cited in the discussion.

Acknowledgments

JCR was funded by the Canadian Frailty Network Interdisciplinary Fellowship and MKB holds a Tier 2 Canada Research Chair in Mobility, Aging and Chronic Disease.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of data: search strategies, full text exclusion reasons, detailed intervention reasons, risk of bias summary, physical activity meta-analyses, and secondary outcome meta-analyses.
[DOC File , 6397 KB - jmir_v24i10e36134_app1.doc]

References

1. World Population Ageing 2017 - Highlights (ST/ESA/SER.A/397). Department of Economic and Social Affairs, United Nations. 2017. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf [accessed 2021-12-09]
2. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
3. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. Geneva, Switzerland: World Health Organization; 2020 Nov 25. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2021-12-09]

4. Gomes M, Figueiredo D, Teixeira L, Poveda V, Paúl C, Santos-Silva A, et al. Physical inactivity among older adults across Europe based on the SHARE database. Age Ageing 2017 Jan 20;46(1):71-77 [FREE Full text] [doi: 10.1093/ageing/afw165] [Medline: 28181637]
5. Kohl 3rd HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, Lancet Physical Activity Series Working Group. The pandemic of physical inactivity: global action for public health. Lancet 2012 Jul 21;380(9838):294-305. [doi: 10.1016/S0140-6736(12)60898-8] [Medline: 22818941]
6. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
7. Cunningham C, O' Sullivan R, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: a systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020 May;30(5):816-827. [doi: 10.1111/sms.13616] [Medline: 32020713]
8. World Health Organization. Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World. Geneva, Switzerland: World Health Organization; 2019.
9. WHO Coronavirus Disease (COVID-19) Dashboard. Secondary WHO Coronavirus Disease (COVID-19) Dashboard 2021. World Health Organization. 2021. URL: https://covid19.who.int [accessed 2021-12-09]
10. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
11. D'Adamo H, Yoshikawa T, Ouslander JG. Coronavirus disease 2019 in geriatrics and long-term care: the ABCDs of COVID-19. J Am Geriatr Soc 2020 May;68(5):912-917. [doi: 10.1111/jgs.16445] [Medline: 32212386]
12. Stockwell S, Trott M, Tully M, Shin J, Barnett Y, Butler L, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. BMJ Open Sport Exerc Med 2021 Feb 1;7(1):e000960 [FREE Full text] [doi: 10.1136/bmjsem-2020-000960] [Medline: 34192010]
13. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
14. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. License: CC BY-NC-SA 3.0 IGO. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/252529 [accessed 2021-12-09]
15. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
16. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019 Aug 30;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]
17. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2021-06-15]
18. Núñez de Arenas-Arroyo S, Cavero-Redondo I, Alvarez-Bueno C, Sequí-Domínguez I, Reina-Gutiérrez S, Martínez-Vizcaíno V. Effect of eHealth to increase physical activity in healthy adults over 55 years: a systematic review and meta-analysis. Scand J Med Sci Sports 2021 Apr;31(4):776-789. [doi: 10.1111/sms.13903] [Medline: 33280182]
19. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
20. Liu JY, Kor PP, Chan CP, Kwan RY, Sze-Ki D. The effectiveness of a wearable activity tracker (WAT)-based intervention to improve physical activity levels in sedentary older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2020 Jul 26;91:104211. [doi: 10.1016/j.archger.2020.104211] [Medline: 32739713]
21. Sohaib Aslam A, van Luenen S, Aslam S, van Bodegom D, Chavannes NH. A systematic review on the use of mHealth to increase physical activity in older people. Clinical eHealth 2020;3:31-39 [FREE Full text] [doi: 10.1016/j.ceh.2020.04.002]
22. Franssen WM, Franssen GH, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
23. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]

24. D'Amore C, Reid JC, Chan M, Fan S, Huang A, Louie J, et al. Smart technology vs. face-to-face physical activity interventions in older adults: a systematic review protocol. JBI Evid Synth 2021 Sep 07;19(10):2801-2812. [doi: 10.11124/JBIES-21-00072] [Medline: 34494613]
25. 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]
26. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
27. de Blok BM, de Greef MH, ten Hacken NH, Sprenger SR, Postema K, Wempe JB. The effects of a lifestyle physical activity counseling program with feedback of a pedometer during pulmonary rehabilitation in patients with COPD: a pilot study. Patient Educ Couns 2006 Apr;61(1):48-55. [doi: 10.1016/j.pec.2005.02.005] [Medline: 16455222]
28. Hansen H, Bieler T, Beyer N, Kallemose T, Wilcke JT, Østergaard LM, et al. Supervised pulmonary tele-rehabilitation versus pulmonary rehabilitation in severe COPD: a randomised multicentre trial. Thorax 2020 May;75(5):413-421 [FREE Full text] [doi: 10.1136/thoraxjnl-2019-214246] [Medline: 32229541]
29. Kawagoshi A, Kiyokawa N, Sugawara K, Takahashi H, Sakata S, Satake M, et al. Effects of low-intensity exercise and home-based pulmonary rehabilitation with pedometer feedback on physical activity in elderly patients with chronic obstructive pulmonary disease. Respir Med 2015 Mar;109(3):364-371 [FREE Full text] [doi: 10.1016/j.rmed.2015.01.008] [Medline: 25682543]
30. Mendoza L, Horta P, Espinoza J, Aguilera M, Balmaceda N, Castro A, et al. Pedometers to enhance physical activity in COPD: a randomised controlled trial. Eur Respir J 2015 Feb;45(2):347-354 [FREE Full text] [doi: 10.1183/09031936.00084514] [Medline: 25261324]
31. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
32. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
33. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-term effectiveness of a comprehensive Internet-based and patient-specific telerehabilitation program with text messaging support for cardiac patients: randomized controlled trial. J Med Internet Res 2015 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
34. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
35. Roberts LM, Jaeger BC, Baptista LC, Harper SA, Gardner AK, Jackson EA, et al. Wearable technology to reduce sedentary behavior and CVD risk in older adults: a pilot randomized clinical trial. Clin Interv Aging 2019 Oct 23;14:1817-1828 [FREE Full text] [doi: 10.2147/CIA.S222655] [Medline: 31695350]
36. Alonso-Domínguez R, Patino-Alonso MC, Sánchez-Aguadero N, García-Ortiz L, Recio-Rodríguez JI, Gómez-Marcos MA. Effect of a multifactorial intervention on the increase in physical activity in subjects with type 2 diabetes mellitus: a randomized clinical trial (EMID Study). Eur J Cardiovasc Nurs 2019 Jun;18(5):399-409. [doi: 10.1177/1474515119835048] [Medline: 30808196]
37. Weinstock RS, Brooks G, Palmas W, Morin PC, Teresi JA, Eimicke JP, et al. Lessened decline in physical activity and impairment of older adults with diabetes with telemedicine and pedometer use: results from the IDEATel study. Age Ageing 2011 Jan;40(1):98-105. [doi: 10.1093/ageing/afq147] [Medline: 21081539]
38. Christiansen MB, Thoma LM, Master H, Voinier D, Schmitt LA, Ziegler ML, et al. Feasibility and preliminary outcomes of a physical therapist-administered physical activity intervention after total knee replacement. Arthritis Care Res (Hoboken) 2020 May;72(5):661-668 [FREE Full text] [doi: 10.1002/acr.23882] [Medline: 30908867]
39. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. J Am Geriatr Soc 2003 Mar;51(3):387-392. [doi: 10.1046/j.1532-5415.2003.51113.x] [Medline: 12588583]
40. Yates T, Davies M, Gorely T, Bull F, Khunti K. Effectiveness of a pragmatic education program designed to promote walking activity in individuals with impaired glucose tolerance: a randomized controlled trial. Diabetes Care 2009 Aug;32(8):1404-1410 [FREE Full text] [doi: 10.2337/dc09-0130] [Medline: 19602539]
41. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth brisk walking intervention on increasing physical activity in older people with cognitive frailty: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16596 [FREE Full text] [doi: 10.2196/16596] [Medline: 32735218]
42. King AC, Campero MI, Sheats JL, Castro Sweet CM, Hauser ME, Garcia D, et al. Effects of counseling by peer human advisors vs computers to increase walking in underserved populations: the COMPASS randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1481-1490 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4143] [Medline: 32986075]

43. Armit CM, Brown WJ, Ritchie CB, Trost SG. Promoting physical activity to older adults: a preliminary evaluation of three general practice-based strategies. J Sci Med Sport 2005 Dec;8(4):446-450. [doi: 10.1016/s1440-2440(05)80060-x] [Medline: 16602173]
44. Audsley S, Kendrick D, Logan P, Jones M, Orton E. A randomised feasibility study assessing an intervention to keep adults physically active after falls management exercise programmes end. Pilot Feasibility Stud 2020 Mar 7;6:37 [FREE Full text] [doi: 10.1186/s40814-020-00570-9] [Medline: 32161660]
45. Mouton A, Cloes M. Efficacy of a Web-based, center-based or combined physical activity intervention among older adults. Health Educ Res 2015 Jun;30(3):422-435. [doi: 10.1093/her/cyv012] [Medline: 25772974]
46. World Report on Ageing and Health. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1&isAllowed=y [accessed 2020-01-22]
47. 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]
48. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ, GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? BMJ 2008 May 03;336(7651):995-998 [FREE Full text] [doi: 10.1136/bmj.39490.551019.BE] [Medline: 18456631]
49. Higgins JP, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ, USA: Wiley Online Library; 2019.
50. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
51. Physical Activity Guidelines for Americans. 2nd edition. U.S. Department of Health and Human Services. 2018. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2021-03-22]
52. Pashazadeh Kan F, Raoofi S, Rafiei S, Khani S, Hosseinifard H, Tajik F, et al. A systematic review of the prevalence of anxiety among the general population during the COVID-19 pandemic. J Affect Disord 2021 Oct 01;293:391-398 [FREE Full text] [doi: 10.1016/j.jad.2021.06.073] [Medline: 34246947]
53. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
54. Yorston LC, Kolt GS, Rosenkranz RR. Physical activity and physical function in older adults: the 45 and up study. J Am Geriatr Soc 2012 Apr;60(4):719-725. [doi: 10.1111/j.1532-5415.2012.03906.x] [Medline: 22486736]
55. McGarrigle L, Todd C. Promotion of physical activity in older people using mHealth and eHealth technologies: rapid review of reviews. J Med Internet Res 2020 Dec 29;22(12):e22201 [FREE Full text] [doi: 10.2196/22201] [Medline: 33372894]
56. Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. Res Synth Methods 2014 Sep;5(3):221-234. [doi: 10.1002/jrsm.1106] [Medline: 26052848]

Abbreviations

COPD: chronic obstructive pulmonary disease
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HRQoL: health-related quality of life
MD: mean difference
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach; submitted 03.01.22; peer-reviewed by A Videira-Silva, S De La Torre; comments to author 19.04.22; revised version received 31.05.22; accepted 09.08.22; published 31.10.22.

Please cite as:
D'Amore C, Reid JC, Chan M, Fan S, Huang A, Louie J, Tran A, Chauvin S, Beauchamp MK
Interventions Including Smart Technology Compared With Face-to-face Physical Activity Interventions in Older Adults: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(10):e36134
URL: https://www.jmir.org/2022/10/e36134
doi:10.2196/36134
PMID:36315229

Commentary

Practical Considerations and Recommendations for "a Revised Hippocratic Oath for the Era of Digital Health"

Rachele Hendricks-Sturrup[1*], MS, MA, DHS; Maryam Nafie[1*], BA; Christine Lu[2,3*], MSc, PhD

[1]Duke-Margolis Center for Health Policy, Washington, DC, United States

[2]Harvard Pilgrim Health Care Institute, Boston, MA, United States

[3]Department of Population Medicine, Harvard Medical School, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Rachele Hendricks-Sturrup, MS, MA, DHS
Duke-Margolis Center for Health Policy
1201 Pennsylvania Avenue NW, 5th Floor
Washington, DC, 20004
United States
Phone: 1 202 621 2800
Email: rachele.hendricks.sturrup@duke.edu

Related Article:

Comment on: http://www.jmir.org/2022/9/e39177/

Abstract

The Hippocratic Oath (the "Oath") is a longstanding body of ethical tenets that have undergone several amendments to accommodate changes and evolutions in the practice of medicine. In their recent perspective entitled, "A Revised Hippocratic Oath for the Era of Digital Health," Meskó and Spiegel offered proposed amendments to the Oath to address both challenges and needs that follow digital health implementation in clinical practice. In this commentary, we offer additional thoughts and considerations to Meskó and Spiegel's proposed amendments to accomplish two goals: (1) reflect on the shared goals and values of all digital health stakeholders and (2) drive home the focus on affirming patient choice, autonomy, and respect.

(J Med Internet Res 2022;24(10):e43383) doi:10.2196/43383

KEYWORDS

digital health; Hippocratic Oath; eHealth; ethics; digital divide

Introduction

History repeatedly reveals that as society faces accelerated change brought by industrial revolution, both institutions and those operating within them must adapt to and endure such changes to survive. Health care institutions, and the practice of medicine in general, are no exception, as we see today in the digital health era.

Meskó and colleagues [1,2] have defined the "digital health era" as today's era in which a "cultural transformation of how disruptive technologies that provide digital and objective data accessible to both health care providers and patients leads to an equal-level doctor-patient relationship with shared decision-making and the democratization of care." Therefore, this definition naturally sparks a multilayered discussion around the ethics of digital health implementation in clinical practice. The longstanding Hippocratic Oath (the "Oath"), for example, and as discussed recently by Meskó and Spiegel [3] in their latest perspective entitled, "A Revised Hippocratic Oath for the Era of Digital Health," is one such level at which the basic ethical tenets of health care can or should be reimagined.

In this commentary, we aim to accomplish two goals in response to Meskó and Spiegel's [3] proposed changes to the Oath: (1) reflect on the shared goals and values of all digital health stakeholders and (2) drive home the focus on affirming patient choice, autonomy, and respect.

Key Considerations and Recommendations

As the medical community contemplates Meskó and Spiegel's [3] proposed new text (in brackets) for the digital health era,

we offer line-by-line comments and considerations that serve to encourage deeper thought around the real-world implications for a potentially revised Oath.

> *I will respect the hard-won scientific gains of those physicians, [researchers, and patients] in whose steps I walk, and gladly share such knowledge as is mine with those who are to follow.*

Today, health care is accelerated by the rapid development and implementation of electronic health records, patient-provider portals, mobile health apps, wearable biosensors, artificial intelligence, social media platforms, etc, in brick-and-mortar, remote, and virtual reality settings. Those responsible for such rapid developments, including their clinical implementation, are clinicians in general (not just physicians), inventors, patients, insurers, technology developers, venture capitalists, and many other stakeholders. Their collective hard-won scientific gains should be acknowledged in the Oath to not just give credit as due but also offer transparency around who is involved in the scientific advancements that drive health care in the 21st century.

> *I will apply, for the benefit of [the healthy and] the sick, all measures [that] are required, avoiding those twin traps of overtreatment and therapeutic nihilism.*

Healthy patients and patients with low health care utilization, for whatever reason, often lack a digital footprint in health care settings (ie, lack an electronic medical record history). Therefore, all patients with seemingly low or lack of health care utilization may erroneously be interpreted by artificial intelligence or machine learning algorithms that process digital health data (eg, electronic medical record data) as "healthy." This is especially true for noncentralized health care systems like those within the United States, where patients may either lack a digital health record altogether due to a lack of insurance status or have fragmented digital health records due to multiple changes in employer-sponsored insurance coverage. This may lead to negative consequences, including inappropriate recommendations for patients based on incorrect estimates of health care utilization patterns. Therefore, it is important to consider potential algorithmic errors that accompany sole or vast reliance on digital health tools in lieu of adopting a more holistic and interpersonal approach to patient care. Additionally, it is important to contemplate the role that digital health plays in triggering illness in seemingly healthy individuals (eg, social media contributing to the onset of depression, anxiety due to overscreening, etc).

> *I will remember that there is an art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife, the chemist's drug [or the programmer's algorithm].*

Patients will increasingly gain digital identities across a growing range of sensitive health care scenarios that may reach beyond any programmer's algorithm (eg, cancer radiology, mental health or substance abuse, family planning, etc). Therefore, we argue that empathy is the utmost imperative for the Oath to ensure that patients are not treated merely as data subjects.

> *[I will treat my patients in an equal-level partnership, and] I will not be ashamed to say "I know not," nor will I fail to call in my colleagues when the skills of another are needed for a patient's recovery.*

Patients are increasingly cost-conscientious and, therefore, have growing needs and demands for cost-related conversations with clinicians. Moreover, equal-level partnerships in patient-provider settings are complicated by histories of systemic racism that have created power imbalances between clinicians and patients. The paternalistic nature of digital health surveillance complicates this matter, making equal-level partnerships potentially illusory to those who have been subjected to negative experiences in the pursuit of health care. Last but not least, an enormous amount of information and power asymmetry exists today between patient communities and health systems, which contributes to health disparities and poor clinical or health outcomes for certain groups of people. Therefore, embedding concepts of equal-level partnership in the Oath may render it infeasible in practice due to long-standing biases and inequities that are deeply rooted in many health care systems everywhere.

> *I will respect the privacy of my patients [and their data], for their problems are not disclosed to me that the world may know.*

A vast amount of digital health data, particularly in the consumer health space and marketplace, fall outside of the scope of existing laws that may protect patient privacy (eg, the US Health Insurance Portability and Accountability Act). In addition, health information privacy is often a matter of context, whereas digital data that are presumably non–health-related can become health-related depending on when, where, why, and by whom the data are collected and used (eg, ridesharing and geolocation apps may collect data about patients' whereabouts around or outside of a medical campus). Furthermore, patients may unknowingly generate data that can become leveraged in the data marketplace or another venue without the patients' consent. Therefore, clinicians should fully consider the privacy practices of digital software or device vendors, health systems, and others to determine whether these proposed changes to the Oath are truly feasible in practice. This is especially given that patients generate large amounts of data as health consumers in general, causing clinicians to rarely encounter or use such vast quantities of data in medical practice.

> *I will remember that I do not treat a fever chart, a cancerous growth, [a data point, or an algorithm's suggestion,] but a human being.*

Patients may knowingly or unknowingly become data subjects. While data are usually averaging a population, clinicians should always focus on the individual patient sitting in front of them. Therefore, today it is critical to create and pave a clear path toward reimagining and reaffirming patient autonomy and respect across all clinical practice areas and settings in which digital health is or may become implemented.

Acknowledging Shared Goals and Patient Choice, Autonomy, and Respect in the Digital Era

Although the Oath was developed in ancient Greece, Meskó and Spiegel [3] noted that the Oath has undergone several amendments, with perhaps the most recent being led by the World Medical Association in 1948, resulting in the Declaration of Geneva [4]. Importantly, the Declaration of Geneva helped drive greater acknowledgment toward shared goals and values among clinical stakeholders, as well as patient autonomy and dignity. These goals are congruent with and complement our goals for this commentary and should, therefore, not be remiss.

Acknowledgments

We would like to acknowledge and thank Dr Christina Silcox at the Duke-Margolis Center for Health Policy for providing initial thoughts and considerations that informed the development of this commentary.

Conflicts of Interest

RH-S reports contract work with the National Alliance Against Disparities in Patient Health.

References

1. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: 10.2196/12490] [Medline: 30912758]
2. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep;3:38-38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
3. Meskó B, Spiegel B. A revised Hippocratic Oath for the era of digital health. J Med Internet Res 2022 Sep 07;24(9):e39177 [FREE Full text] [doi: 10.2196/39177] [Medline: 36069845]
4. WMA Declaration of Geneva. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-geneva/ [accessed 2022-10-08]

Edited by T Leung; submitted 10.10.22; this is a non–peer-reviewed article;accepted 11.10.22; published 26.10.22.

Please cite as:
Hendricks-Sturrup R, Nafie M, Lu C
Practical Considerations and Recommendations for "a Revised Hippocratic Oath for the Era of Digital Health"
J Med Internet Res 2022;24(10):e43383
URL: https://www.jmir.org/2022/10/e43383
doi:10.2196/43383
PMID:36287597

Viewpoint

The Limitations of User-and Human-Centered Design in an eHealth Context and How to Move Beyond Them

Lex van Velsen[1,2]; Geke Ludden[3], PhD; Christiane Grünloh[1,4], PhD

[1]eHealth Department, Roessingh Research and Development, Enschede, Netherlands

[2]Department of Communication Science, University of Twente, Enschede, Netherlands

[3]Department of Design, Production and Management, University of Twente, Enschede, Netherlands

[4]Biomedical Systems and Signals group, University of Twente, Enschede, Netherlands

Corresponding Author:
Geke Ludden, PhD
Department of Design, Production and Management
University of Twente
Drienerlolaan 5
Enschede, 7522NB
Netherlands
Phone: 31 534892920
Email: g.d.s.ludden@utwente.nl

Abstract

Human-centered design (HCD) is widely regarded as the best design approach for creating eHealth innovations that align with end users' needs, wishes, and context and has the potential to impact health care. However, critical reflections on applying HCD within the context of eHealth are lacking. Applying a critical eye to the use of HCD approaches within eHealth, we present and discuss 9 limitations that the current practices of HCD in eHealth innovation often carry. The limitations identified range from limited reach and bias to narrow contextual and temporal focus. Design teams should carefully consider if, how, and when they should involve end users and other stakeholders in the design process and how they can combine their insights with existing knowledge and design skills. Finally, we discuss how a more critical perspective on using HCD in eHealth innovation can move the field forward and offer 3 directions of inspiration to improve our design practices: value-sensitive design, citizen science, and more-than-human design. Although value-sensitive design approaches offer a solution to some of the biased or limited views of traditional HCD approaches, combining a citizen science approach with design inspiration and imagining new futures could widen our view on eHealth innovation. Finally, a more-than-human design approach will allow eHealth solutions to care for both people and the environment. These directions can be seen as starting points that invite and support the field of eHealth innovation to do better and to try and develop more inclusive, fair, and valuable eHealth innovations that will have an impact on health and care.

(J Med Internet Res 2022;24(10):e37341) doi:10.2196/37341

KEYWORDS

user-centered design; human-centered design; eHealth, value-sensitive design; citizen science

Introduction

Background

For years, user-centered design (UCD) has been considered to be a crucial part of eHealth design. It is believed to improve an innovation's usefulness and usability [1], to improve end-user satisfaction with the innovation [2], and to increase the quality of user requirements [3]. In addition to this, in the context of eHealth, a strong focus on end users during the design process (patients, care professionals, or others) is deemed to improve adoption rates [4-6], patient decision-making [7], patient engagement [8,9], and patient satisfaction [8]. UCD is a design approach or philosophy that originated in the 1980s. Two seminal publications coined the concept [10] and listed its key principles [11]. According to Gould and Lewis [11], these key principles are that there should be an early focus on users and tasks. First, designers should study the users and the tasks that they need to perform with a technology to understand them fully. Second, the design team should use empirical measurements. Prospective end users need to work with prototypical versions of a technology and their performance and reactions should then be analyzed in a scientific manner. Third, one should apply iterative design. Within the design process, there should be multiple cycles of design, testing, and redesign. Since then, different publications have provided

hands-on guidelines on how to implement UCD in practice, for example, an overview by Maguire [12] of the complete UCD process and methods to apply at every stage, and an international standard offering guidance on human-centered design (HCD) activities (International Organization for Standardization 9241-210 "Human-centred design for interactive systems," the latest version being from 2019 [13]). According to the International Organization for Standardization standard, HCD "is an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors or ergonomics and usability knowledge and techniques." The international standard outlines the following principles that should be followed in a human-centered approach [13]:

1. The design is based upon explicit understanding of users, tasks, and environments.
2. Users are involved throughout design and development.
3. The design is driven and refined by user-centered evaluation.
4. The process is iterative.
5. The design addresses the whole user experience.
6. The design team includes multidisciplinary skills and perspectives.

Meanwhile, in design research as well as in the field of human-computer interaction (HCI), the term UCD has become a topic of much debate. Although the terms UCD and HCD are often used interchangeably, several authors in these fields have argued that the term UCD reduces a person to someone using a technology, failing to see the whole human being that lives with the technology. Gasson [14], for example, argues for using HCD instead of UCD to avoid a focus on people as technology users and to allow a broader view of human activity supported by technology. In this paper, we will use the term HCD as an umbrella term for both approaches.

HCD has found its way into eHealth design via several road maps that specifically focus on health and well-being as a domain. The Center for eHealth Research (CeHReS) road map [15] is perhaps the most widely used design road map in the field and specifies the following 5 main phases in which HCD is a crucial element: contextual inquiry, value specification, design, operationalization, and summative evaluation. Each phase comes with its own goals and selection of methods that one can apply. The CeHReS road map has been used to guide the development of a wide range of eHealth apps, such as a mobile app to support people in dealing with ticks and tick bites [16], an information dashboard to support nurses in antimicrobial stewardship [17], or a blended exercise therapy intervention for patients with knee and hip osteoarthritis [18]. Other design approaches that heavily hinge on the HCD philosophy and have been used to guide eHealth development include intervention mapping [19], the person-based approach [20], and Integrate, Design, Assess, and Share [21]. Although the authors of these different approaches have all defined their own phases in the design process, their makeup and essence are basically the same. They revolve around extensive end-user (and stakeholder) involvement, iterative design, and working systematically, the same principles that the founders of HCD listed.

Many articles report on the results of HCD processes in eHealth and reflect on the experiences of research and design teams while using the approach (eg, the studies by Fico et al [4], Kramer et al [22], and Atkinson et al [23]). This has led to a body of literature in which the approach receives a lot of praise, with authors claiming that applying the approach has improved the quality of their eHealth service. However, critiquing one's own approach too much would decrease the value of the design process results as well as the value of the resulting publication. This publication bias may have led the community to believe that HCD is, by definition, the best approach toward eHealth innovation. In addition, this belief in HCD as the best approach has influenced the writers of calls in funding programs (such as the European Union's Horizon program), in which HCD is often included as a prerequisite for funding. Focusing calls in such a way pushes researchers to use an approach that may not be optimal for their context [24]. In all, it is crucial that the eHealth design community acknowledges the inherent limitations of HCD in eHealth, to (1) reduce the positive bias in their reflections on the application of HCD, (2) raise awareness among design teams about the limitations of this design approach, and (3) improve HCD processes for eHealth by accounting for the limitations of the approach.

Objectives

In this paper, we aimed to provide an overview of the limitations of using HCD in an eHealth context. These limitations were derived from our own experiences in numerous HCD processes for eHealth services as well as from the larger body of human-centered eHealth design studies. We would like to clarify that we are not opponents of HCD. A simple Google Scholar search of our publication records will show that we have used the approach in the past [17,25,26] and have reflected on its merits. To elaborate on some of our experiences, the second author (GL) and colleagues in the MinD—Designing for people with dementia project reflected on their experiences and on the complexities of involving people with dementia in the design and evaluation of (digital) tools that could improve their psychosocial well-being [27,28]. In another project, we involved children with breathing problems in the design of a smart wearable [29]. Both projects aimed to design for a group of people that was very different from the project team, which called for the inclusion of end-user experiences in the design process. The publications referenced here explain how the teams benefited from working with these groups, creating end results that were (more) acceptable. Nevertheless, in both cases, the teams clearly faced challenges when it came to selection of participants and representation of the complete target group as well as the interpretation of data gathered during cocreation sessions. There were discussions in which the information or knowledge of experts by experience conflicted with related work and knowledge of the project team. In the smart wearable project, we started with including the children who would eventually use the tool in the design process, although child pulmonologists were also involved. A much larger and more diverse group of stakeholders, including child physiotherapists, is involved in a currently running follow-up to this project. This greatly adds to the complexity of decision-making, as was also confirmed in another study in which the third author (CG) was

involved. This study concluded that when multiple groups of stakeholders are involved, more knowledge is needed on how to deal with conflicting perspectives [30].

Although we are convinced that HCD approaches are necessary in the design of eHealth innovations, we think that it would be healthy for the community if critical reflection on using HCD becomes common practice, with the ultimate goal of improving our use of HCD. Of course, we are not the first (nor will we be the last) to critically reflect on the concept of HCD. Therefore, before listing the limitations that we would like to stress, we summarize some of the critical reflections on HCD that have been published in the past.

Previous Reflections on HCD

In his 2005 essay, "Human-Centered Design Considered Harmful," Norman [31] reflects on some of the principles that underlie HCD; for example, the principle that technology should always adapt to people and not the other way around. He posits that this principle is not really true as people indeed adapt to technology; moreover, technology changes and continues to change our behaviors and our lives. In fact, in eHealth innovation, providing a health intervention that changes people's behavior or their perspective on health and even their lives is often the aim. According to Norman [31], improving some aspects for individuals or groups may worsen them for others, and the focus on humans and their needs distracts from other design-related activities and may lead to incoherent and complex designs. Norman [31], therefore, suggested an activity-centered design approach, which includes a deep understanding of people but also fosters a deep understanding of technology, tools, and the reasons for the activities.

In 2011, a study by Bannon [32] revisited the roots of the HCI discipline to argue that HCI should develop an even more human-centered approach. For instance, a focus on augmenting people's existing skills goes beyond merely *considering the user* (and their requirements) and instead, also prioritizes the understanding of people, their concerns, activities, and, in particular, their values and more fundamental needs when designing new technology [32]. Forlizzi [33] urges moving beyond UCD toward stakeholder-centered and service design. Similar to Bannon [32], Forlizzi [33] also reflects on how drastically the field of HCI and also technology and society have evolved, broadening the focus of HCI from ergonomics and usability to also include experience, engagement, and entertainment. Forlizzi [33] identified the lack of an economic perspective in UCD approaches; this perspective is needed, given that today technologies are increasingly being designed as services used by multiple stakeholders. Hence, she urges the HCI community to move beyond UCD and to consider a service design approach that also includes economics [33]. More recently, the top-down approach of HCD as being traditionally led by professionals has been criticized, inviting the exploration and discussion of *community-driven design* [34,35]. The argument goes that today's global challenges deal with complex sociotechnical systems that require a bottom-up approach in which communities themselves take the lead to solve problems collaboratively, facilitated by professionals [34,36]. The value of actively involving citizens and communities as coresearchers is also well known in approaches such as participatory action research and citizen science [37-39].

These reflections on HCD show that the primary focus on "knowing the end user" is too narrow, and extensions or new approaches have been developed to respond to the need for including additional perspectives and dimensions. Although it incorporates some of the views mentioned previously, this paper specifically focuses on health and well-being as an application domain and the limitations of HCD that we experienced and observed in previous eHealth projects.

Limitations of HCD in eHealth

Limitation 1: HCD Tends to Lead to Sampling Bias

"It's not for me, but my neighbor would love this" is what we often hear participants say during design sessions. It makes one wonder, if we always hear this, are we designing something that nobody wants? Or are we talking to the wrong people? HCD methods most often rely on studying a relatively small sample in depth; this approach is prone to a range of sampling biases. By default, we are not talking to the right people, as not everybody can join the sessions; some specific groups will not be represented at all, whereas other groups are overrepresented (a selection bias) [40]. For example, in a case study by Haslwanter et al [41] that aimed to design a product that enables older adults to stay independently at home for longer, certain methods such as inviting the target audience to a demonstration house led to a recruitment bias in terms of gender, level of mobility, and interest in technology. This sampling bias is further compounded by the difficulty of finding people who are willing and able to spend their time in an interview or design activity. Participating in a design study can be quite time-consuming, meaning that patients need to combine this with their (high) disease burden. Jongsma and Friesen [42] pointed out that participatory research is either too demanding and therefore unfeasible or too uninclusive and therefore unfair. In other words, you sometimes have to take what you get at the cost of biasing your sample (self-selection bias). This bias has been made explicit for experimental research in HCI, which tends to be biased toward younger, more tech-savvy, and more educated participants [43]. For design activities, it is not clear what characterizes those that are more willing to join design activities. In addition, it is also possible, especially among professional participants, that one person participates on behalf of a department or professional group. This person may or may not be chosen by management and acts as a barrier toward involving other people within that group (gatekeeper bias). One could argue that these biases are not a large issue, as qualitative research does not strive toward generalizability, but if we want to include a diverse range of views and contexts, we should strive toward some form of generalizability [44], and we must therefore take these biases seriously. Finally, when biased end-user input is the only source of inspiration for a new eHealth intervention, the implication is that the biased design input is translated into a biased technology. Many authors acknowledge this bias in the Limitations sections of their articles but do not discuss how this bias affected the design or in what way they tried to negate this bias in the design or evaluation phase. Could

it be that the design research community has just accepted this bias and its consequences as facts of life?

We would advise design researchers to go further than just naming sampling bias as a limitation in their publication and being done with it. In addition, mapping exactly what the bias looks like (ie, who the most important excluded groups are) and stating how the neglected groups will be included in the design process (eg, in an additional round of design activities using a method that is especially suited for these groups) or in the evaluation activities (eg, targeting the sample of a prototype evaluation toward these groups) would be a great improvement.

Limitation 2: End-User Input Might Be Biased and Limited

The premise of HCD is that listening to end users and incorporating their needs and wishes into eHealth design will ensure an innovation that end users will want and can use. However, listening to end users (patients or care professionals) has its limitations. First, a lot of the knowledge, opinions, or attitudes that are crucial for eHealth service design are tacit. Tacit knowledge is developed from direct experience and action and is highly pragmatic, and subconsciously understood [45]. Tacit knowledge has been found to be of paramount importance for care professionals for developing working routines [46,47]. Similarly, patient end users (in either preventive or curative care) will have internalized routines and assumptions that they rely on. The challenge with tacit knowledge is that it is difficult to verbalize, and thus elicit, during HCD [48]. Most HCD studies tend to overly rely on traditional interviews and focus groups. These methods are well suited to pose direct questions but are therefore limited in their capacity to elicit tacit knowledge (unless combined with other methods, such as observations). To solve this issue, more creative methods should be applied that have the power to elicit tacit knowledge indirectly, or that allow for the researcher to determine tacit knowledge or procedures for themselves. Two such methods are narrative inquiry [49] and the critical decision method [50]. Next, patients and care professionals apply work-arounds to get things or to get their work done [51,52]. Although this breaking out of protocol might be considered undesirable, it might also be necessary to achieve the best possible outcome for a patient. For example, Yang et al [53] describe the case of a hospital information system that recommends medication dosages, in which physicians override the system as it does not properly take into account pediatric dosages. Furthermore, Dannecker et al [54] describe a myriad of work-arounds that patients with osteoarthritis apply in order to construct pain intensity ratings. On the one hand, work-arounds are a fantastic source of inspiration for design. On the other hand, they can also be a problem as patients or care professionals may not want to disclose them because they are breaking the rules or protocols while implementing them. Zheng et al [55] provide an overview of these challenges and ways to overcome them.

Limitation 3: HCD Tends to Lead to Overreliance on (Fresh) End-User Input

Overview

Consulting the end user early on and throughout the development process of an eHealth service is an important principle of HCD. However, end-user input (or stakeholder input, for that matter; Limitation 4: End users Are Only a Subset of the People Who Should Be Heard During eHealth Design section) is not necessarily the only source of input for designers. The actual design of a service is a creative process that can be fueled by *end-user input* (which can be translated into requirements), but it can also be served by the knowledge and skills of a (multidisciplinary) design team and its creativity, or a by *technology push*.

End-User Input

In many articles that describe the HCD process of an eHealth service, end-user input seems to outweigh all the other elements that should inform a thorough design process. We stated in the Introduction section that HCD is part of many (if not most) current eHealth innovation projects. This has created a body of knowledge on user needs and requirements, but this body of knowledge is rarely used to inform other projects. Instead, every project runs its own interview, focus group, and design sessions and the secondary use of end-user input is disregarded. There are some recent exceptions in the field of designing for dementia that resulted in design tools that can, for some part, take over the contextual inquiry phase (such as the MinD toolkit [56]). These largely evolved because of the difficulties of working with this end-user group, but similar tools might be helpful for other patient groups and could well reduce the burden on patients. In addition to this, many designers of eHealth tools are involved in a series of eHealth innovations, in many cases making them experienced “understanders” of particular patient groups. Creativity, so it is argued, is a valuable means of design for solving ill-defined problems. It can drive both the instrumental and hedonic aspects of an innovation in terms of functionality, safety, usability, and affect [57]. A thorough understanding of the implementation context and of end users’ needs and wishes is paramount for creative design [58], and thus, HCD can play a very valuable role in preparing the stage for creative thinking.

Technology push seems to have become a dirty word in the eHealth community, and user input, translated into requirements, has become the driving force in many eHealth development processes. This preference for end-user input over propagating technological innovations touches upon the classic debate of the technology push versus the demand pull. Can end users imagine what they actually want, or can they only repeat what they have already seen? Even if people are encouraged to imagine functions that they would like to see in a product or service, this may then relate to *imagined* needs instead of actual user needs. This imposes the risk that these functions are not used in the future [41]. Furthermore, what is the most successful innovation strategy, developing what the market wants, or creating what is technologically possible? Although there are different camps in the scientific community with regard to this issue, it seems like technology push and demand pull are

dependent on each other for developing a valuable and successful innovation. Although a technology push is often considered to be the core source of innovation, a demand pull can also drive innovation by bringing forth new ideas and concepts from the users and their context and is always necessary for ensuring economic viability [59]. It seems that, rather than figuring out which approach is best, we should investigate how both approaches can be combined [60].

In line with current open science approaches, design (research) teams should make more efforts to make end-user input reusable and to reuse it where possible. Moreover, rather than thinking that end-user input is the only source of inspiration that can lead to value-adding eHealth innovations, they should aim to find the sweet spot where end user consultation, technology push, and the design team's knowledge and creativity coincide or come together right. Or, as Norman [31] put it, "Paradoxically, the best way to satisfy users is sometimes to ignore them."

Limitation 4: End Users Are Only a Subset of the People Who Should Be Heard During eHealth Design

End users of eHealth solutions are most often patients, care professionals, and citizens, and are crucial in terms of being taken into account when designing a new eHealth service. However, there are also other organizations and actors involved, that is, all stakeholders. Stakeholders can be classified as being either direct or indirect [61]. Direct stakeholders are individuals or organizations who interact directly with the system, whereas indirect stakeholders are affected by the use of the system. Although indirect stakeholders do not interact with the technology themselves, they can exert influence over an innovation or experience consequences from its implementation and use. For example, patients as indirect stakeholders are often not considered when developing electronic medical records, even though the records are about them [62]. Similarly, in the case of developing an eHealth service that gives patients access to their records in Sweden, the "medical profession was not really perceived as a legitimate actor in the development process" [63]. At the same time, the medical profession (clinicians and nurses) in this particular case contested the very idea of the project and was not interested in participating in the design process [64], which created another challenge.

It is imperative to consult both end users and other stakeholders in order to develop a service model and a business model. A service model is an overview of how a technological service interacts with end users and stakeholders, as well as with any other services (on the web or offline). As such, it is a combination of the patient journey and the care protocol (or care path) envisioned for an eHealth service [64]. A business model, on the other hand, is an overview of how the eHealth service is being brought to the market and how it is envisioned to sustain. Both the service and the business model are of paramount importance for creating an eHealth service that is durable and that will be accepted beyond its end users.

Holistic design, in which end users (primary and secondary), lead users, and other parties that can exert influence over the implementation and success of an eHealth service are involved, and in which technology, a service model, and a business model are developed simultaneously, is increasingly being used as a successor to HCD. It is at the core of the CeHReS road map, and in recent years, many developers and researchers have reported their experiences with holistic design (and the CeHReS road map) in case studies (eg, the study by van Velsen et al [25]). Holistic design, in its turn, has disadvantages and challenges that design teams will have to deal with (such as ensuring the collaboration of health professionals [65] and ensuring proper expectation management among all stakeholders [66]). The involvement of different stakeholder groups can provide challenges in terms of balancing their influence as well as the potential assumptions that user groups have of each other that might be based on stereotypes [41,67].

Limitation 5: Understanding the Added Value of HCD Is Complicated

In general, HCD is considered to be a valuable approach that results in better eHealth services, the point of reference here being eHealth services that are developed without user involvement in any form. This is supported by literature claiming, for example, that user involvement was positive overall and through intermediate factors such as better user requirements [68] or by the revised version of a website based on user input being preferred [7]. A systematic mapping study showed that user participation and involvement can have a positive effect on system success (eg, user satisfaction, ease of use, and system use), but it has also been shown to have negative correlations with system success in older studies [2]. It has been acknowledged that measuring user participation is complex; there is no common conceptual model to measure and validate this effect [2] and we do not have a complete understanding of how user involvement affects product development [68]. In practice, the conclusion that HCD leads to better eHealth services is made through a subjective reflection on the design process by the authors of an article describing the design process. Owing to the competitive nature of academia and the need to publish (or perish), researchers are subject to a (subliminal) bias and are prone to being overly positive about their results [69]. This may mean that our general opinion about HCD is based on a large body of subjective viewpoints.

So, how can we make an objective assessment of the value of HCD for the design of eHealth innovations? If we would fall back on the traditional means to assess the quality of an intervention in health care, then the logical thing to do would be to create a single design briefing, give it to one design team that will apply HCD and another design team that will apply an alternative design process that does not include end users directly and evaluate the resulting eHealth services in terms of usefulness, innovativeness, and usability. This way, we would be able to compare whether one approach "performs" better than the other. However, such studies are difficult to perform (as one has to duplicate the design process) and comparing one "condition" with the other is difficult, as there is also creativity and skill involved in design. Controlling for creativity and skill within the comparison among design approaches would be challenging and maybe even impossible. Despite all the challenges involved in setting up a fair comparison between HCD and an alternative approach, Guo et al [70] conducted a survey study among 389 Chinese digital start-ups and found that applying either a customer orientation or a technology

orientation could lead to successful business models. However, combining the 2 approaches led to troublesome situations, as resources were limited, and it was difficult to combine the business logics involved.

The easiest solution toward "unbiasing" our understanding of the value of HCD in the eHealth context would be to adopt a critical view toward the value of the approach in different case studies even if this comes at the cost of a critical peer review. For example, Kip et al [71] critically reflected on their design activities for developing a virtual reality application for practicing coping skills for clients in forensic mental health care and provided an overview of the suitability of different HCD methods for the target population (including both successes and failures). Only by publishing our failures and critical reflections can we create a proper and more nuanced view on the value of HCD (methods) and help eHealth innovation to mature as a research field.

Limitation 6: HCD Risks of Supporting the Status Quo

When it comes to developing innovative and disruptive eHealth services, questions posed to prospective end users are naturally going to be hypothetical. So, responses are likely be limited to end users' ability to envision new concepts (More Than Needs and Wishes section). Although it is the role of the designer to develop new concepts, when we try to understand people's needs, this is more easily done in the current context and not in the future context for which the design is to be developed. Take the example of developing a technology for the prediction of exacerbations based on real-world data and using these predictions for shared decision-making between patients and professionals. This future scenario is so far removed from the current care setting that it is difficult for people to reflect and articulate associated needs. Sometimes, the future scenario is not very far in the future, but the imagined needs do not necessarily reflect the actual needs when it comes to that specific situation. This was illustrated in the study by Haslwanter et al [41] in which there was a difference in what people wanted while seeing the demonstration house and what people then used when implemented in practice. The well-known colloquialism, "what people say they do vs what they actually do" comes to mind, which conveys what we argue to be an even greater challenge when it comes to future scenarios.

Although this limitation might be overcome by not relying solely on end-user input, there is also the risk that incorporating end-user input might hamper innovation. Indeed, it might lead to concluding that the status quo is the most desirable future. For example, research related to patients reading their electronic health records showed that health care professionals question the abilities of patients to understand these records and voice concerns which do not necessarily materialize [30,72]. Here, the desired status quo was a cumbersome process with patients asking for permission to access their paper-based records. It might seem trivial that one stakeholder group is not able to assess the needs of another group; however, health care professionals as domain experts are often considered to be an authority when designing eHealth systems [30]. Similarly, ageism (whereby people hold prejudices about older adults) is often used to think about older adults' use and ability to use new technology. How aging is framed (eg, as a "problem" to be managed by technology [73]) can also represent common stereotypes and limit design opportunities. As a result, design teams are reluctant to introduce new (technological) concepts, as the general opinion is that older adults do not want change and are unable to deal with new innovations. However, when Jung and Ludden [74,75] interviewed older adults with mobility impairments and presented them with the prospect of using exoskeleton technology, a technology that they were completely unfamiliar with, they seemed rather open to this possible future. But generally, it seems easier for people (patients or health care professionals) to imagine barriers than to imagine opportunities for developing a new type of care, working routine, or society. Consequently, in HCD, we have the tendency to design something new for the current world, rather than designing a new world.

In order to move beyond our prejudices and current (working) routines, there are several things we can do, but these require us to change *how* we do HCD. In their discussion of designing against the status quo, Khovanskaya et al [76] offered several pieces of advice. Designers will need to study and understand the history behind the current situation and the prejudices therein. Then, in order to envision a new reality, designers might need to resort to different sources of inspiration, besides end-user input, such as feminist and queer theory, art, or the maker culture. The trick for the design team will then be to introduce these (disruptive) new ideas to potential end users and stakeholders and to create a safe space in which these ideas can be presented and discussed. Designing against the status quo might mean designing for the long term. The health care setting is conservative and reluctant to change. Therefore, combining short-term ambitions and design ideas (closer to the status quo) with long-term ambitions and design ideas (closer to the disruptive vision) is an approach that is most likely to succeed.

Limitation 7: Traditional HCD and Designing for Behavior Change Are Not a Good Match

With the increasing importance of preventing chronic diseases and improving lifestyle in general, many eHealth services aim to change the behavior of end users. They must support people to quit smoking, to sit less, or to eat healthier. This trend has led to a research discipline called persuasive design or design for behavior change. Persuasive design is concerned with developing technology "to reinforce, change or shape attitudes or behaviors, or both, without using coercion or deception" [77] and has been found to increase compliance with eHealth services [78]. In design for behavior change, a range of tools and methods have been developed with specific attention to the eHealth context [79]. Although from a normative standpoint, persuading people to perform certain health-improving behaviors might be desirable, it does infringe on the person's autonomy. For example, a study in the context of smoking cessation showed that although a person might want to stop smoking, they still might not want to make a commitment to behavior change [80]. In a discussion on the ethics of persuasive design; therefore, Berdichevsky and Neuenschwander [81] posit one golden rule for persuasive design, which is as follows: "The creators of a persuasive technology should never seek to persuade a person

or persons of something they themselves would not consent to be persuaded to do." However, the rise of monitoring and coaching technology and the need to make the population adopt a healthier lifestyle have created a situation in which many technologies are being developed that aim to persuade people to adopt a certain behavior *eventually*, while also applying an HCD approach. In this case, however, it is impossible to question potential end users (through interviews, focus groups, and design sessions) about this future goal. Their *initial* standpoint toward a change in behavior may be negative, although at the same time they may have a positive attitude toward caring about (and monitoring) their health. For example, one can probe how one should persuade or support patients with diabetes to be more physically active, but if the participant is unmotivated to do so (eg, the participant is perfectly happy with their current lifestyle), every question or probe is likely to result in a negative reply, if not an aversion to the design session in itself, or could lead to a socially accepted reply (not reflecting the participant's attitude) just to be over and done with the session. The problem here is that the (technological) solution direction of the design team conflicts with the person's wishes, desires, or values.

In short, persuasive design and HCD seem to form an unhappy marriage. Therefore, if one were to design a technology that aims to induce health behavior change, one might best trade 1 of the 2 in for something else. Instead of persuasive design, one could resort to using *tuning* as a paradigm that focuses on building internal self-knowledge and self-awareness by supporting appropriate knowledge, skills, and practice [82]. Instead of a single-factor health guidance (eg, to walk 10,000 steps a day), this approach acknowledges the complexities of health in terms of an individual's context and other behaviors and aims to "support a person gaining knowledge, skills, and practice of how to tune their health across contexts" [82]. In addition, taking into account the end user's stage of chance (following the transtheoretical model [83]) in eHealth design and personalization will ensure that content, functionalities, and design strategies [84] are geared toward the aspects of behavior change to which the end user is most receptive. If one is quite attached to persuasive design, one could trade HCD in for value-sensitive design (VSD). Rather than focusing on what persuasive technology should do and how (as one would do in HCD), VSD aims to understand why a design might be harmful, and it will reveal the value conflicts or tensions that must be solved [85]. The latter approach will respect the participants and their context and will not evoke negative emotions. Once the value conflicts are fully mapped, it will be the design team's task to create a design that is capable of reaching the behavior change goals while respecting the end users' values. Or, one could go even one step further and supply VSD with capability sensitive design [86]. In such an approach, the design team has to elicit not only *what* end users value but also whether these outcomes *ought* to be valued.

Limitation 8: HCD Tends to Miss Out on Ethical, Societal, and Political Aspects

HCD activities focus on individual users, their context, and their needs and expectations in relation to specific tasks and goals. Thus, HCD tends to prioritize the microlevel rather than the mesolevel and macrolevel. However, organizational aspects on the mesolevel are crucial when it comes to the implementation of eHealth solutions in real life (End users Are Only a Subset of the People Who Should Be Heard During eHealth Design section). HCD supports the economic and social pillars of sustainability [13]. However, ethical, societal, and political issues on the macrolevel can be overlooked when focusing on the individual user.

Technological advancements such as machine learning and artificial intelligence (AI), have the potential to support people in their everyday tasks (eg, decision support tools in health care). However, they also have the potential to increase inequality by amplifying biases and assumptions that are invisible to users. This has been outlined in the book "Weapons of Math Destruction," where mathematical models and algorithms are typed as *opaque* (lacking transparency or completely invisible), *damaging* (harmful or unfair for certain people and creating pernicious feedback loops), or that *scale* (have the capacity to grow exponentially) [87]. Negative examples include decision support for judges using recidivism models or a university ranking model that creates an ecosystem of education and industry of tutors that adapt to that scoring system [87]. As more and more decisions are automated in the future based on AI, algorithmic biases potentially lead to discrimination based on certain characteristics such as income, education, gender, or ethnicity [88]. Although these decisions may work for many, people who fall outside what has been incorporated in the design as the "norm" (eg, through specific training data or through how the models are designed) have to deal with the consequences without the opportunity to appeal the decisions. Discrimination may stay hidden if we focus on specific user groups and how technology can support their tasks and goals (eg, supporting a judge in sentencing decisions).

These examples show that the human perspective is often not considered or only very narrowly considered for a small user group. The focus on the individual user within an HCD can be complemented by approaches such as VSD, responsible research and innovation [89], or human-centered explainable AI [90], which aim to incorporate ethical and societal aspects into the design process. Within HCD alone, this is difficult to address, especially if the innovations are so complex that it is difficult to communicate the risks or form a black box by definition. Furthermore, as COVID-19 pandemic measures and tracking apps showed, certain decisions include not only a technological, scientific, and societal perspective but also a political one [91,92]. Given the circumstances, political decisions might prioritize certain values over others (eg, public security over privacy in relation to track and trace), this issue is also relevant for eHealth technologies (eg, apps used for contact tracing and risk information [93]).

Limitation 9: HCD Thinks About the Beginning but Not the End

Most eHealth services that are developed through an HCD approach are accompanied by elaborate onboarding procedures and implementation plans. The desire to reach and secure a high number of end users makes sense, as one of the common key performance indicators for these services is the number of end

users served (for a longer period). Interestingly, this focus on the first use of the service seems to come at a cost. It rarely happens that a design process also devotes attention and time to longer-term use or to ending the use of a service. Should an app change at some point in the user's journey? When has an eHealth service fulfilled its purpose? How do we determine this moment? Which actions are associated with ending the end-user journey supported by the eHealth service? It rarely happens that answers to these questions are sought and processed into service design for the eHealth context.

A topic that is associated with ending the personal use of the eHealth service is ending the eHealth service completely. It may feel a bit contradictory to think about the terminating of a service during the design stage, but for some services, this will be crucial for acceptance and for preventing undesirable situations. The COVID-19 pandemic has made the need for these deimplementation plans very clear. Contact tracing apps were built on top of the privacy-preserving exposure notification frameworks developed by Google and Apple. Despite the fact that these frameworks did not require tracking the geographical location of end users, they were met with a lot of skepticism and privacy concerns. Indeed, one cannot exclude the possibility that tracking the geographical location by means of this technology might be possible in the future. So although the use of these technologies might be legitimate and useful for the short term, they might be harmful in the long run. Therefore, the introduction of eHealth services that come with large implications, such as the COVID-19 contact tracing apps, should be accompanied by a plan that specifies when we can stop using them and how we can erase all the data that they collect during their lifetime.

Discussion

Nine Limitations

In this paper, we have described 9 limitations that we currently see with the application of HCD in eHealth. This set of limitations came about by critically reflecting on our own eHealth innovation projects and by reviewing the body of work in this domain. Of course, not all limitations are restricted to the eHealth context and many of them are applicable to the full range of digital services one can develop. However, we felt that it was important to provide a complete overview of the main limitations that we have seen. Again, we would like to emphasize that, even after composing this list of limitations, we do feel that HCD processes have their place in the design of eHealth innovations, especially in combination with other sources of input, such as available knowledge and a technology push. Our objective was to provide a wake-up call to researchers and designers in the eHealth domain. Although some actively seek and implement ways to improve the role of HCD in their innovations, others continue to rely on standard (and suboptimal) ways to involve end users. The most important point we want to make is that critical reflection on applying HCD methods in the design of eHealth services is lacking, and this is not helping the field of eHealth innovation to mature. This list of limitations is most probably not conclusive, and we hope that more critical reflection by other researchers in the field will eventually lead to a better understanding of (1) how and when HCD methods can really contribute to the design process (and when they would not do so); (2) how HCD methods can lead to more generalizable knowledge that the field needs and could share; and (3) how HCD methods can be integrated in the process of multidisciplinary design teams that include relevant health, technological, ethical, and other expertise.

In addition to the limitations, this paper also discusses several current developments and opportunities to improve the use of HCD in the design of eHealth innovations. This means that there are already signs that the field of eHealth innovation is changing and taking important next steps that change how we see and deal with HCD in eHealth innovation. We briefly want to reflect on and elaborate on 3 of these developments here as we see them as very important for the future of eHealth innovation.

More Than Needs and Wishes

In limitations 4, 7, and 8, we have mentioned how VSD or value-based design and also capability sensitive design can provide guidance in involving multiple stakeholders and integrating ethical perspectives in the design process. VSD defines human values as "what is important to people in their lives, with a focus on ethics and morality" [94]. Although VSD displays some similarities to HCD, it includes aspects that go beyond it as well, such as the commitment to analyze both direct and indirect stakeholders; to distinguish designer values, stakeholder values, and values explicitly supported by the project; to conduct an analysis on individual, group, and societal levels; and the possibility for technology and social structures to coevolve [94]. Hence, VSD may offer a solution to some of the biased or limited views of traditional HCD approaches. For instance, the commitment of a thorough analysis of direct and indirect stakeholders might mitigate the risk of sampling bias (limitation 1) and bias toward and overreliance on end-user input (limitations 2 and 3), as it opens up the design space in terms of who should get a seat at the table. The challenge in applying VSD is that a focus on values can lead to a rather abstract understanding of what end users and other stakeholders consider important. This means that it leaves a serious task for the design team to translate this understanding into a tangible design, a task in which they may well want to involve end users again. We need a good body of work describing and reflecting on the processes used to do this (eg, the studies by Boerema et al [95] and Smits et al [96]).

More Than Consulting End Users

In limitation 7, we discussed the problem that in some eHealth development processes, the (technological) solution direction of the design team conflicts with a person's wishes, desires, or values. Limitation 7 goes on to discuss ways to deal with this situation, but it also triggers the question of whether the solution direction that the design team was aiming for was the right one. Were they trying to answer the wrong question all along? An approach that tries to tackle this is citizen science, which seeks to engage "citizens" (ie, everyone who at some point may deal with the outcomes of science) in research in different ways. Citizen science has been around for a while now, and its uptake and importance in health and biomedical research are growing

[97]. Citizen science overlaps with HCD in its methods and aims and can range from *contributory* (eg, participation of the public or patients through data collection and processing) to *collaborative* (eg, public involvement in refining research questions, analyzing data, or disseminating findings), *cocreation* (eg, researchers and members of the public working together across key research processes), and *extreme citizen science* (in which researchers provide tools and methods to enable communities to develop their own participatory research projects) [38]. In the health and well-being domain, the people dealing with the outcome of research are often also the subject of the research (patients or experts by experience). A larger adoption of citizen science could, on the one hand, help us in making the transition from seeing patients as subjects whose opinions we politely ask for, to coresearchers who are active in not only providing data or answers to our questions but also in asking the right questions and setting a research agenda for (public) health. On the other hand, citizen science philosophy somewhat conflicts with the issues and recommendations that we mention in limitation 6 (*HCD Risks to Support the Status Quo*). As we discuss in limitation 6, in order to envision a new reality, designers or a design team should bring in inspiration, come up with new ideas and imagine new futures, and find ways to introduce and discuss these with the public or the community they are working with. The 10 principles of citizen science, put forward by the European Citizen Science Association in 2015 [98], provide an initial set of guidelines to take up citizen science in eHealth innovation. It is up to the field of eHealth innovation to further discuss and critically reflect on how and when to use citizen science approaches in the eHealth context.

More Than Humans

In limitation 8, we discuss how HCD risks missing out on ethical, societal, and political aspects. A prioritization of the microlevel, the personal level, has consequences, especially when this is the preferred and largely dominant approach in a particular field. One such consequence could be that we disregard the impact that our innovations have on the environment. A recent appeal in *The Lancet Digital Health* stressed this very point [99]. In addition, several researchers have called for moving beyond the dominant anthropocentric perspective to include nonhuman perspectives [100,101]. As Giaccardi and Redström [101] describe, we may at some point reach the boundaries of what can be conceived through UCD and HCD processes. The increasing complexity of what we can design and the increasing consequences that our designs can have call for new methods and for reconsideration of the role of HCD methods and the weight given to them.

Conclusions

In this paper, we have presented 9 limitations of using HCD in eHealth and 3 directions of inspiration to improve design practices. We feel that these directions provide good starting points to do better and to try and develop more inclusive, fair, and valuable eHealth innovations that will have an impact on health and care. We trust and hope that this discussion of limitations as well as this short outlook to the future of eHealth innovation will stir up a bit of dust and would be very happy to see others add to it so that with more careful, considered, and critical use of HCD we can improve our eHealth research and innovation methods together.

Acknowledgments

This work has been supported by the TOPFIT Citizenlab project, funded by the Regio Deal Twente and the Agenda voor Twente.

Conflicts of Interest

None declared.

References

1. Vredenburg K, Mao J, Smith P, Carey T. A survey of user-centered design practice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2002 Presented at: CHI02: Human Factors in Computing Systems; Apr 20 - 25, 2002; Minneapolis Minnesota USA. [doi: 10.1145/503376.503460]
2. Abelein U, Paech B. Understanding the influence of user participation and involvement on system success – a systematic mapping study. Empir Software Eng 2013 Dec 05;20(1):28-81. [doi: 10.1007/s10664-013-9278-4]
3. Kujala1 S. Effective user involvement in product development by improving the analysis of user needs. Behav Inform Technol 2008 Nov 17;27(6):457-473. [doi: 10.1080/01449290601111051]
4. Fico G, Martinez-Millana A, Leuteritz J, Fioravanti A, Beltrán-Jaunsarás ME, Traver V, et al. User centered design to improve information exchange in diabetes care through ehealth : results from a small scale exploratory study. J Med Syst 2019 Nov 18;44(1):2. [doi: 10.1007/s10916-019-1472-5] [Medline: 31741069]
5. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, et al. A user-centered design and usability testing of a web-based medication reconciliation application integrated in an eHealth network. Int J Med Inform 2019 Jun;126:138-146. [doi: 10.1016/j.ijmedinf.2019.03.013] [Medline: 31029255]
6. Mitzner T, Dijkstra K. Evaluating User Centered Design of E-Health for Older Adults: Using Subjective Methods. USA: IGI Global; 2016.
7. Atkinson NL, Massett HA, Mylks C, McCormack LA, Kish-Doto J, Hesse BW, et al. Assessing the impact of user-centered research on a clinical trial eHealth tool via counterbalanced research design. J Am Med Inform Assoc 2011 Jan 01;18(1):24-31 [FREE Full text] [doi: 10.1136/jamia.2010.006122] [Medline: 21169619]

8. Nazi KM, Turvey CL, Klein DM, Hogan TP. A decade of veteran voices: examining patient portal enhancements through the lens of user-centered design. J Med Internet Res 2018 Jul 10;20(7):e10413 [FREE Full text] [doi: 10.2196/10413] [Medline: 29991468]
9. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
10. Norman D, Draper S. User Centered System Design New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.
11. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985 Mar;28(3):300-311. [doi: 10.1145/3166.3170]
12. Maguire M. Methods to support human-centred design. Int J Human Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
13. ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: human-centred design for interactive systems. ISO. 2019 Jul. URL: https://www.iso.org/standard/77520.html [accessed 2022-03-10]
14. Gasson S. Human-centered vs. user-centered approaches to information system design. J Inform Technol Theory Application 2003 Jan;5(2):29-46.
15. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
16. van Velsen L, Beaujean DJ, Wentzel J, Van Steenbergen JE, van Gemert-Pijnen JE. Developing requirements for a mobile app to support citizens in dealing with ticks and tick bites via end-user profiling. Health Informatics J 2015 Mar;21(1):24-35 [FREE Full text] [doi: 10.1177/1460458213496419] [Medline: 24282189]
17. Wentzel J, van Velsen L, van Limburg M, de Jong N, Karreman J, Hendrix R, et al. Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med Inform Decis Mak 2014 Jun 05;14(1):45 [FREE Full text] [doi: 10.1186/1472-6947-14-45] [Medline: 24898694]
18. Bossen D, Kloek C, Snippe HW, Dekker J, de Bakker D, Veenhof C. A blended intervention for patients with knee and hip osteoarthritis in the physical therapy practice: development and a pilot study. JMIR Res Protoc 2016 Feb 24;5(1):e32 [FREE Full text] [doi: 10.2196/resprot.5049] [Medline: 26912378]
19. Boekhout JM, Peels DA, Berendsen BA, Bolman CA, Lechner L. An eHealth intervention to promote physical activity and social network of single, chronically impaired older adults: adaptation of an existing intervention using intervention mapping. JMIR Res Protoc 2017 Nov 23;6(11):e230 [FREE Full text] [doi: 10.2196/resprot.8093] [Medline: 29170146]
20. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
21. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
22. Kramer L, Blok M, van Velsen L, Mulder B, de Vet E. Supporting eating behaviour of community-dwelling older adults: co-design of an embodied conversational agent. Design Health (Abingdon) 2021;5(1):120-139 [FREE Full text] [doi: 10.1080/24735132.2021.1885592] [Medline: 34381936]
23. Atkinson NL, Saperstein SL, Desmond SM, Gold RS, Billing AS, Tian J. Rural eHealth nutrition education for limited-income families: an iterative and user-centered design approach. J Med Internet Res 2009 Jun 22;11(2):e21 [FREE Full text] [doi: 10.2196/jmir.1148] [Medline: 19632974]
24. Hallewell Haslwanter JD, Fitzpatrick G, Miesenberger K. Key factors in the engineering process for systems for aging in place contributing to low usability and success. J Enabling Technol 2018 Oct 10;12(4):186-196. [doi: 10.1108/jet-12-2017-0053]
25. van Velsen L, Illario M, Jansen-Kosterink S, Crola C, Di Somma C, Colao A, et al. A community-based, technology-supported health service for detecting and preventing frailty among older adults: a participatory design development process. J Aging Res 2015;2015:216084 [FREE Full text] [doi: 10.1155/2015/216084] [Medline: 26346580]
26. Hooglugt F, Ludden GD. A mobile app adopting an identity focus to promote physical activity (MoveDaily): iterative design study. JMIR Mhealth Uhealth 2020 Jun 15;8(6):e16720 [FREE Full text] [doi: 10.2196/16720] [Medline: 32538795]
27. Lim JN, Almeida R, Holthoff-Detto V, Ludden GD, Smith T, Niedderer K, MinD consortium. What is needed to obtain informed consent and monitor capacity for a successful study involving People with Mild Dementia? Our experience in a multi-centre study. In: Designing With and for People With Dementia: Wellbeing, Empowerment and Happiness. Dresden: TUD press; 2019.
28. Niedderer K, Holthoff-Detto V, van Rompay TJ, Karahanoğlu A, Ludden GD, Almeida R, et al. This is Me: evaluation of a boardgame to promote social engagement, wellbeing and agency in people with dementia through mindful life-storytelling. J Aging Stud 2022 Mar;60:100995 [FREE Full text] [doi: 10.1016/j.jaging.2021.100995] [Medline: 35248311]

29. Siering L, Ludden GD, Mader A, van Rees H. A theoretical framework and conceptual design for engaging children in therapy at home-the design of a wearable breathing trainer. J Pers Med 2019 May 20;9(2):27 [FREE Full text] [doi: 10.3390/jpm9020027] [Medline: 31137523]
30. Cajander A, Grünloh C. Electronic health records are more than a work tool: conflicting needs of direct and indirect stakeholders. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK.
31. Norman DA. Human-centered design considered harmful. Interact 2005 Jul 01;12(4):14-19. [doi: 10.1145/1070960.1070976]
32. Bannon L. Reimagining HCI. Interact 2011 Jul 01;18(4):50-57. [doi: 10.1145/1978822.1978833]
33. Forlizzi J. Moving beyond user-centered design. Interactions 2018 Aug 22;25(5):22-23. [doi: 10.1145/3239558]
34. Hekler E, Taylor J, Dow S, Morris M, Grant F, Phatak S, et al. Proceedings of the Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion. 2019 Presented at: DIS '19: Designing Interactive Systems Conference 2019; Jun 23 - 28, 2019; San Diego CA USA. [doi: 10.1145/3301019.3319998]
35. Norman D, Spencer E. Community-based, human-centered design. jnd.org. 2019 Jan 3. URL: https://jnd.org/community-based-human-centered-design/ [accessed 2021-08-18]
36. Meyer M, Norman D. Changing design education for the 21st Century. She Ji J Design Econ Innov 2020;6(1):13-49 [FREE Full text] [doi: 10.1016/j.sheji.2019.12.002]
37. Wilson B. A short history of community-driven design. In: Resilience for All. Washington, DC: Island Press; 2018.
38. Borda A, Gray K, Downie L. Citizen science models in health research: an Australian commentary. Online J Public Health Inform 2019;11(3):e23 [FREE Full text] [doi: 10.5210/ojphi.v11i3.10358] [Medline: 31993110]
39. Oberschmidt K, Grünloh C, Nijboer F, van Velsen L. Best practices and lessons learned for action research in eHealth design and implementation: literature review. J Med Internet Res 2022 Jan 28;24(1):e31795 [FREE Full text] [doi: 10.2196/31795] [Medline: 35089158]
40. Latulipe C, Gatto A, Nguyen H, Miller D, Quandt S, Bertoni A, et al. Design considerations for patient portal adoption by low-income, older adults. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:3859-3868 [FREE Full text] [doi: 10.1145/2702123.2702392] [Medline: 27077140]
41. Hallewell Haslwanter JD, Fitzpatrick G. Why do few assistive technology systems make it to market? The case of the HandyHelper project. Univ Access Inf Soc 2016 Oct 21;16(3):755-773. [doi: 10.1007/s10209-016-0499-3]
42. Jongsma K, Friesen P. The challenge of demandingness in citizen science and participatory research. Am J Bioeth 2019 Aug 24;19(8):33-35. [doi: 10.1080/15265161.2019.1619867] [Medline: 31544651]
43. Dickinson A, Arnott J, Prior S. Methods for human – computer interaction research with older people. Behav Inform Technol 2007 Jul;26(4):343-352. [doi: 10.1080/01449290601176948]
44. Groger L, Mayberry PS, Straker JK. What we didn't learn because of who would not talk to us. Qual Health Res 2016 Jul 01;9(6):829-835. [doi: 10.1177/104973299129122180]
45. McAdam R, Mason B, McCrory J. Exploring the dichotomies within the tacit knowledge literature: towards a process of tacit knowing in organizations. J Knowl Manag 2007;11(2):43-59. [doi: 10.1108/13673270710738906]
46. Thornton T. Tacit knowledge as the unifying factor in evidence based medicine and clinical judgement. Philos Ethics Humanit Med 2006 Mar 17;1(1):E2 [FREE Full text] [doi: 10.1186/1747-5341-1-2] [Medline: 16759426]
47. Kothari AR, Bickford JJ, Edwards N, Dobbins MJ, Meyer M. Uncovering tacit knowledge: a pilot study to broaden the concept of knowledge in knowledge translation. BMC Health Serv Res 2011 Aug 18;11(1):198 [FREE Full text] [doi: 10.1186/1472-6963-11-198] [Medline: 21851617]
48. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
49. Kothari A, Rudman D, Dobbins M, Rouse M, Sibbald S, Edwards N. The use of tacit and explicit knowledge in public health: a qualitative study. Implement Sci 2012 Mar 20;7:20 [FREE Full text] [doi: 10.1186/1748-5908-7-20] [Medline: 22433980]
50. Okechukwu OJ, Weller G, Watt J. Eliciting experts? knowledge in emergency response organizations. Int J Emergency Services 2014;3(2):118-130. [doi: 10.1108/IJES-06-2014-0009]
51. Bjørn P, Balka E. Health care categories have politics too: unpacking the managerial agendas of electronic triage systems. In: Proceedings of the Tenth European Conference on Computer Supported Cooperative Work. 2007 Presented at: Tenth European Conference on Computer Supported Cooperative Work; Sep 24-28, 2007; Limerick, Ireland.
52. Fitzpatrick G, Ellingsen G. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. Comput Supported Coop Work 2012 Jun 21;22(4-6):609-665. [doi: 10.1007/s10606-012-9168-0]
53. Yang Z, Ng B, Kankanhalli A, Luen Yip JW. Workarounds in the use of IS in healthcare: a case study of an electronic medication administration system. Int J Human Comput Stud 2012 Jan;70(1):43-65 [FREE Full text] [doi: 10.1016/j.ijhcs.2011.08.002]
54. Dannecker E, Warne-Griggs M, Royse L, Hoffman K. Listening to patients' voices: workarounds patients use to construct pain intensity ratings. Qual Health Res 2019 Mar;29(4):484-497. [doi: 10.1177/1049732318773714] [Medline: 29890886]

55. Zheng K, Ratwani RM, Adler-Milstein J. Studying workflow and workarounds in electronic health record–supported work to improve health system performance. Ann Intern Med 2020 Jun 02;172(11_Supplement):S116-S122. [doi: 10.7326/m19-0871]
56. MinD tools - Designing for people with dementia homepage. MinD tools - Designing for people with dementia. URL: https://designingfordementia.eu/resources/mind-tools [accessed 2022-06-06]
57. Zeng L, Proctor R, Salvendy G. Creativity in ergonomic design: a supplemental value-adding source for product and service development. Hum Factors 2010 Aug;52(4):503-525. [doi: 10.1177/0018720810376056] [Medline: 21141243]
58. Christiaans H, Venselaar K. Creativity in design engineering and the role of knowledge: modelling the expert. Int J Technol Des Educ 2005 Jan;15(3):217-236. [doi: 10.1007/s10798-004-1904-4]
59. Di Stefano G, Gambardella A, Verona G. Technology push and demand pull perspectives in innovation studies: current findings and future research directions. Res Policy 2012 Oct;41(8):1283-1295 [FREE Full text] [doi: 10.1016/j.respol.2012.03.021]
60. Brem A, Voigt K. Integration of market pull and technology push in the corporate front end and innovation management—insights from the German software industry. Technovation 2009 May;29(5):351-367 [FREE Full text] [doi: 10.1016/j.technovation.2008.06.003]
61. Friedman B, Kahn P, Borning A, Huldtgren A. Value sensitive design and information systems. In: Early engagement and new technologies: Opening up the laboratory. Dordrecht: Springer; 2013.
62. Friedman B, Kahn JP, Borning A. Value sensitive design and information systems. In: Human-computer Interaction Management Information Systems: Foundations. Armonk, NY: ME Sharpe; 2006.
63. Erlingsdóttir G, Lindholm C. When patient empowerment encounters professional autonomy: the conflict and negotiation process of inscribing an eHealth service. Scandinavian J Public Administration 2015;19(2):48.
64. Broekhuis M, Weering MD, Schuit C, Schürz S, van Velsen L. Designing a stakeholder-inclusive service model for an eHealth service to support older adults in an active and social life. BMC Health Serv Res 2021 Jul 05;21(1):654 [FREE Full text] [doi: 10.1186/s12913-021-06597-9] [Medline: 34225745]
65. Noergaard B, Sandvei M, Rottmann N, Johannessen H, Wiil U, Schmidt T, et al. Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR Res Protoc 2017 May 17;6(5):e75 [FREE Full text] [doi: 10.2196/resprot.7084] [Medline: 28526674]
66. van Dooren MM, Siriaraya P, Visch V, Spijkerman R, Bijkerk L. Reflections on the design, implementation, and adoption of a gamified eHealth application in youth mental healthcare. Entertainment Comput 2019 Aug;31:100305 [FREE Full text] [doi: 10.1016/j.entcom.2019.100305]
67. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" patient participation through eHealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
68. Kujala S. User involvement: a review of the benefits and challenges. Behav Inform Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
69. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US States data. PLoS One 2010 Apr 21;5(4):e10271 [FREE Full text] [doi: 10.1371/journal.pone.0010271] [Medline: 20422014]
70. Guo H, Wang C, Su Z, Wang D. Technology push or market pull? Strategic orientation in business model design and digital start‐up performance*. J Prod Innov Manag 2020 May 12;37(4):352-372 [FREE Full text] [doi: 10.1111/jpim.12526]
71. Kip H, Kelders SM, Bouman YH, van Gemert-Pijnen LJ. The importance of systematically reporting and reflecting on eHealth development: participatory development process of a virtual reality application for forensic mental health care. J Med Internet Res 2019 Aug 19;21(8):e12972 [FREE Full text] [doi: 10.2196/12972] [Medline: 31429415]
72. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
73. Vines J, Pritchard G, Wright P, Olivier P, Brittain K. An age-old problem. ACM Trans Comput Hum Interact 2015 Mar 04;22(1):1-27. [doi: 10.1145/2696867]
74. Jung M, Ludden G. Potential of exoskeleton technology to assist older adults with daily living. In: Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3170427.3188464]
75. Jung MM, Ludden GD. What do older adults and clinicians think about traditional mobility aids and exoskeleton technology? J Hum Robot Interact 2019 Jun 20;8(2):1-17. [doi: 10.1145/3311789]
76. Khovanskaya V, Dombrowski L, Harmon E, Korn M, Light A, Stewart M, et al. Designing against the status quo. interactions 2018 Feb 23;25(2):64-67. [doi: 10.1145/3178560]
77. Oinas-Kukkonen H, Harjumaa M. Towards deeper understanding of persuasion in software and information systems. In: Proceedings of the First International Conference on Advances in Computer-Human Interaction. 2008 Presented at: First International Conference on Advances in Computer-Human Interaction; Feb 10-15, 2008; Sainte Luce, Martinique, France. [doi: 10.1109/ACHI.2008.31]

78. 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]
79. Niedderer K, Clune S, Ludden G, editors. Design for Behaviour Change Theories and Practices of Designing for Change. Abingdon, UK: Routledge; 2017.
80. Ploderer B, Smith W, Howard S, Pearce J, Borland R. Things you don't want to know about yourself: ambivalence about tracking and sharing personal information for behaviour change. In: Proceedings of the 24th Australian Computer-Human Interaction Conference. 2012 Presented at: OzCHI '12: The 24th Australian Computer-Human Interaction Conference; Nov 26 - 30, 2012; Melbourne Australia. [doi: 10.1145/2414536.2414610]
81. Berdichevsky D, Neuenschwander E. Toward an ethics of persuasive technology. Commun ACM 1999 May;42(5):51-58. [doi: 10.1145/301353.301410]
82. schraefel M, Hekler E. Tuning. interactions 2020 Feb 25;27(2):48-53. [doi: 10.1145/3381897]
83. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Sep;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
84. Ludden G, Hekkert P. Design for healthy behavior: design interventions and stages of change. In: Proceedings of the 9th International Conference on Design and Emotion. 2014 Presented at: 9th International Conference on Design and Emotion; Oct 6-10, 2014; Bogota, Colombia.
85. Davis J. Design methods for ethical persuasive computing. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive 2009; 4th International Conference on Persuasive Technology Claremont; Apr 26 - 29, 2009; California USA. [doi: 10.1145/1541948.1541957]
86. Jacobs N. Capability sensitive design for health and wellbeing technologies. Sci Eng Ethics 2020 Dec 18;26(6):3363-3391 [FREE Full text] [doi: 10.1007/s11948-020-00275-5] [Medline: 33206283]
87. O'Neil C. Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy. New York, United States: Penguin Books Limited; 2016.
88. Ferrer X, Nuenen TV, Such JM, Cote M, Criado N. Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technol Soc Mag 2021 Jun;40(2):72-80. [doi: 10.1109/mts.2021.3056293]
89. Owen R, Macnaghten P, Stilgoe J. Responsible research and innovation: from science in society to science for society, with society. Sci Public Policy 2012 Dec 05;39(6):751-760. [doi: 10.1093/scipol/scs093]
90. Ehsan U, Riedl M. Human-centered explainable AI: towards a reflective sociotechnical approach. In: Proceedings of the HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020. 2020 Presented at: HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020,; Jul 19-24, 2020; Copenhagen, Denmark.
91. Greenhalgh T, Ozbilgin M, Contandriopoulos D. Orthodoxy, illusio, and playing the scientific game: a Bourdieusian analysis of infection control science in the COVID-19 pandemic. Wellcome Open Res 2021 Sep 22;6:126. [doi: 10.12688/wellcomeopenres.16855.2]
92. Meyer J, Kay J, Epstein D, Eslambolchilar P, Tang L. A life of data: characteristics and challenges of very long term self-tracking for health and wellness. ACM Trans Comput Healthcare 2020 Apr 30;1(2):1-4. [doi: 10.1145/3373719]
93. Dubov A, Shoptawb S. The value and ethics of using technology to contain the COVID-19 epidemic. Am J Bioeth 2020 Jul 18;20(7):W7-11 [FREE Full text] [doi: 10.1080/15265161.2020.1764136] [Medline: 32420817]
94. Friedman B, Hendry D. Value Sensitive Design Shaping Technology with Moral Imagination. Cambridge, MA: MIT Press; 2019.
95. Boerema S, van Velsen L, Vollenbroek-Hutten M, Hermens H. Value-based design for the elderly: an application in the field of mobility aids. Assist Technol 2017;29(2):76-84. [doi: 10.1080/10400435.2016.1208303] [Medline: 27532436]
96. Smits M, Ludden G, Peters R, Bredie S, van GH. Values that matter: a new method to design and assess moral mediation of technology. MIT Press Direct 2022;38(1):39-54. [doi: 10.1162/desi_a_00669]
97. Wiggins A, Wilbanks J. The rise of citizen science in health and biomedical research. Am J Bioeth 2019 Aug 24;19(8):3-14. [doi: 10.1080/15265161.2019.1619859] [Medline: 31339831]
98. Gold M. ECSA 10 Principles of citizen science. OSF Home. 2019 Jan 23. URL: https://osf.io/xpr2n/ [accessed 2022-07-25]
99. Chevance G, Hekler EB, Efoui-Hess M, Godino J, Golaszewski N, Gualtieri L, et al. Digital health at the age of the Anthropocene. Lancet Digital Health 2020 Jun;2(6):e290-e291 [FREE Full text] [doi: 10.1016/S2589-7500(20)30130-8] [Medline: 33328121]
100. Coulton P, Lindley J. More-than human centred design: considering other things. Design J 2019 May 29;22(4):463-481. [doi: 10.1080/14606925.2019.1614320]
101. Giaccardi E, Redström J. Technology and more-than-human design. Design Issues 2020 Sep;36(4):33-44. [doi: 10.1162/desi_a_00612]

Abbreviations

AI: artificial intelligence

CeHRes: Center for eHealth Research
HCD: human-centered design
HCI: human-computer interaction
UCD: user-centered design
VSD: value-sensitive design

Edited by T Leung; submitted 16.02.22; peer-reviewed by P Gorp, S Kujala, T Risling; comments to author 03.05.22; revised version received 27.07.22; accepted 19.08.22; published 05.10.22.

Please cite as:
van Velsen L, Ludden G, Grünloh C
The Limitations of User-and Human-Centered Design in an eHealth Context and How to Move Beyond Them
J Med Internet Res 2022;24(10):e37341
URL: https://www.jmir.org/2022/10/e37341
doi:10.2196/37341
PMID:36197718

Viewpoint

Bridging the European Data Sharing Divide in Genomic Science

Fruzsina Molnár-Gábor[1], JD, PhD; Michael J S Beauvais[2,3], BA, BA (Jurisprudence), BCL, JD, MSc; Alexander Bernier[2], BCL, JD, LLM; Maria Pilar Nicolas Jimenez[4], JD, PhD; Mikel Recuero[4], LLB, JD, LLM; Bartha Maria Knoppers[2], BA, LLB, BCL, MA, DLS, PhD

[1]Faculty of Law, University of Heidelberg, Heidelberg, Germany

[2]Centre of Genomics and Policy, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada

[3]Faculty of Law, University of Toronto, Toronto, ON, Canada

[4]Research group in social and legal sciences applied to the new technosciences IT 1541-22, Faculty of Law, University of the Basque Country, Basque University System, Bilbao, Spain

Corresponding Author:
Fruzsina Molnár-Gábor, JD, PhD
Faculty of Law
University of Heidelberg
BioQuant Centre
Im Neuenheimer Feld 267 (BQ267)
Heidelberg, 69120
Germany
Phone: 49 62215451 ext 336
Email: fruzsina.molnar-gabor@uni-heidelberg.de

Abstract

In this viewpoint, we argue for the importance of creating data spaces for genomic research that are detached from contexts in which fundamental rights concerns related to surveillance measures override a purpose-specific balancing of fundamental rights. Genomic research relies on molecular and phenotypic data, on comparing findings within large data sets, on searchable metadata, and on translating research results into a clinical setting. These methods require sensitive genetic and health data to be shared across borders. International data sharing between the European Union (EU) or the European Economic Area and third countries has accordingly become a cornerstone of genomics. The EU General Data Protection Regulation contains rules that accord privileged status to data processing for research purposes to ensure that strict data protection requirements do not impede biomedical research. However, the General Data Protection Regulation rules applicable to international transfers of data accord no such preferential treatment to international data transfers made in the research context. The rules that govern the international transfer of data create considerable barriers to international data sharing because of the cost-intensive procedural and substantive compliance burdens that they impose. For certain jurisdictions and select use cases, there exist practically no lawful mechanisms to enable the international transfer of data because of concerns about the protection of fundamental rights. The proposed solutions further fail to address the need to share large data sets of local and regional cohorts across national borders to enable joint analyses. The European Health Data Space is an emerging federated, EU-wide data infrastructure that is intended to function as an infrastructure bringing together EU health data to improve patient care and enable the secondary use of health-related data for research purposes. Such infrastructure is implementing new institutions to support its functioning and is being implemented in reliance on a new enabling law, the regulation on the European Health Data Space. This innovation provides the opportunity to facilitate EU contribution to international genomic research efforts. The draft regulation for this data space provides for a concept of data infrastructure intended to enable cross-border data exchange and access, including access to genetic and health data for scientific analysis purposes. The draft regulation also provides for obligations of national actors aimed at making data widely available. This effort is laudable. However, in the absence of further, more fundamental changes to the manner in which the EU regulates the secondary use of health data, it is reasonable to believe that EU participation in international genomic research efforts will remain impeded.

(J Med Internet Res 2022;24(10):e37236) doi:10.2196/37236

KEYWORDS

international data transfer; scientific research; genomics; safe data spaces; data protection

The General Data Protection Regulation Hurdle for International Genomic Research

The launch of many large-scale multinational research projects over the past 2 decades exemplifies the importance of international data sharing in genomics and omics research [1,2]. Cross-matching data between centers, establishing large community reference data collections, and accessing external reference data sets enhances the understanding of human biology and disease and benefits translational stratified medicine.

Thus, data protection issues related to international data sharing are inextricably linked to genomic research. Furthermore, where data protection issues are considered in an international context but with European Union (EU) involvement, reference to the EU General Data Protection Regulation (GDPR) [3] becomes unavoidable.

This regulation is perceived to hinder rather than promote cross-border data sharing at the international level [4]. The main objective of the GDPR is to create nuanced rules that balance the benefits and risks of processing personal data and protecting individual interests. The international data transfer mechanisms of the GDPR act to ensure that the obligations applicable according to EU data protection law continue to be applied after the data are transferred outside the EU and the European Economic Area, including the proportionate balancing of benefits and harms.

Regulators and courts often deem unlawful outbound transfers of data from the EU and European Economic Area to third countries that implement considerable State surveillance measures. Such legal determinations act as a functional bar to the outbound transfer of genomic data of EU and European Economic Area provenance to such third countries. Furthermore, such a legal determination places on prospective data transferors the considerable burden of assessing the State surveillance practices of third countries before performing international data transfers in favor thereof. This heightens the complexities and compliance costs inherent in performing outbound data transfers from the EU and the European Economic Area to third jurisdictions.

Thus, any concerns about fundamental rights arising in a third country in relation to surveillance activities carried out by public authorities influence the assessment of the level of protection in that country and lead to its rules being found disproportionate or even disregarding the core of the fundamental rights concerned. This practice precludes outbound transfers of genomic data from being performed unless significant political changes are made to the policing and surveillance practices of those third countries [5]. Although it is crucial to raise the level of data protection in relation to the surveillance activities of State actors from a fundamental rights perspective, this is a long-term endeavor. Until this is achieved, data sharing for socially significant purposes such as scientific research will decline. Removal of these considerable barriers to data sharing in genomic research is contingent on political determinations that are outside the scope of activities of scientific research communities.

In this paper, we argue for the importance of creating data spaces to enable international collaboration in the use of genomic data. These data spaces should be detached from contexts in which fundamental rights concerns related to surveillance measures override a purpose-specific balancing of fundamental rights and of the benefits and risks of processing personal data and protecting individual interests. First, we assess the relevant provisions of the GDPR and detail their implications for data exporters and importers. Second, we outline how State surveillance practices can affect the potential for researchers to share genomic data. Third, we address the fundamental rights context of scientific research. Fourth, we analyze possible solutions that would enable genomic data to be transferred to researchers in countries outside the EU and the European Economic Area without international consensus being achieved on issues of State surveillance. Contractual issues are first discussed, followed by secure data spaces. This structure allows us to address the challenges of international data transfers in the current legal situation in detail, identify the main problems, and, on this basis, consider which solutions could provide a remedy, also in the context of further clarification of the cornerstones of the European Health Data Space (EHDS).

Background: GDPR Transfer Rules

In this section, we describe in considerable detail the international data transfer rules of the GDPR. This discussion provides the necessary context in framing the challenges that the international data transfer rules create for genomic researchers. This description is necessary to demonstrate how our proposed solution responds to the needs of scientific researchers and would also meet the demands of EU and European Economic Area data protection regulators.

The main legal mechanism by which personal data may be transferred from the EU and the European Economic Area to a third country for scientific research purposes is a decision of the European Commission confirming the adequacy of the level of data protection in the recipient country.

The European Commission is responsible for determining whether a third jurisdiction is adequate. Once a jurisdiction has been deemed adequate, outbound transfers of data to that third jurisdiction can be made without additional legal compliance efforts being required. If the destination of an international data transfer is not subject to an adequacy decision, additional measures enabling legal compliance must be implemented before the outbound transfer is performed.

The following criterion is used to determine whether a third jurisdiction can be deemed *adequate*. An *adequate level of data protection* requires that the third country ensure, by virtue of its domestic legislation and international commitments, a level of data protection “essentially equivalent” to that guaranteed in the EU [6].

This does not mean that an identical level of protection is required. The methods used to protect data by the third country may differ from those used in the EU, but such methods must

nevertheless prove *effective in practice* [7]. The GDPR also defines rules guiding the assessment of adequacy as to whether the essence of the fundamental right to data protection is respected and whether its limitation is subject to the principles of necessity and proportionality.

The text of the applicable laws in the concerned jurisdiction is not the sole criterion assessed in performing this evaluation. Indeed, the practices of authorities, administrative bodies, and courts in the country of destination are of equal relevance in assessing whether adequacy status can be lawfully conferred. Decisions of the European Commission establishing private sector areas in the United States as adequate for the purpose of receiving data transfers from the EU have been annulled twice by the Court of Justice of the EU (CJEU). This court is the principal court that is responsible for interpreting the fundamental right to data protection in the EU [5,6]. The annulment of these decisions was rooted in concerns about the fundamental rights to data protection, respect for private life, and effective remedies (ie, the availability of redress mechanisms for affected parties) as defined by the EU Charter of Fundamental Rights, a legally binding catalog of human rights applicable in the EU (cf Article 7: Respect for private and family life, Article 8: Protection of personal data, and Article 47: Right to an effective remedy and to a fair trial). In this decision, the CJEU concluded that the surveillance practices of US authorities and the lack of redress available to EU citizens relative thereto violated the aforementioned human rights guarantees [8].

Therefore, in summary, the European Commission is responsible for ascribing *adequacy* status to countries outside the EU and the European Economic Area. This requires the European Commission to determine that the concerned jurisdiction provides data protection guarantees *essentially equivalent* to those available in the EU.

Recent adequacy decisions portend a change from a relatively lenient adequacy analysis to a more stringent evaluation that requires the legislation and administrative practices in the concerned jurisdiction to mirror those in the EU. This is seen very clearly in the case of Japan [9]. In Japan, the adequacy decision of the European Commission related only to the private sector, considering that oversight mechanisms in data protection law differ in their design in the private and public sectors. Such restrictions on the scope of adequacy can be understood as a strong indicator that sector-specific evaluations of foreign data protection legislation will, in the future, be used to confer adequacy status on a sector-specific or statute-specific basis rather than on a national basis. Given the rigor and granularity of recent European Commission adequacy analyses, it is not surprising that the level of data protection is currently confirmed by an adequacy decision of the European Commission in only 13 countries and territories around the world [10].

Implications of the GDPR Transfer Mechanisms for Data Importers and Exporters

If data are transferred from the EU or the European Economic Area to a jurisdiction that is not subject to an adequacy decision, a distinct legal mechanism must be implemented to ensure the lawfulness of the data transfer. The level of data protection in the third country must be examined not only in the case of an adequacy decision but also when the international data transfer is based on another transfer mechanism established in the GDPR [5]. The categories of available transfer mechanisms include the imposition of additional safeguards that maintain the standard of data protection in the EU and the European Economic Area, which are exhaustively enumerated in the text of the GDPR. At the time of writing, the only transfer safeguard that has been developed in a standardized manner are the Standard Contractual Clauses. These Standard Contractual Clauses require transferors to integrate standard-form contractual language into their data transfer agreements, which imposes numerous GDPR-derived compliance-related obligations on the recipients of the data transfer, effectively approximating the extraterritorial application of the GDPR. The other transfer mechanism is a context-specific derogation from the application of the GDPR to enable a specific transfer that could not otherwise be performed in compliance with the requirements of the GDPR. There is a considerable range of derogations that are potentially relevant to international transfers of genomic data. However, these derogations are intended to enable personal data to be transferred out of the EU or the European Economic Area on an ad hoc, exceptional basis. Therefore, Standard Contractual Clauses are the only real mechanism that could at present enable continuous, ongoing international transfers of personal data to jurisdictions that do not benefit from an adequacy decision.

Senders and recipients of personal data who base their data exchange on contractual clauses such as the Standard Contractual Clauses of the European Commission are also obliged to verify before each transfer whether the level of data protection secured under EU law is met in the recipient country. Accordingly, the Standard Contractual Clauses can only be drawn on to secure international data transfers if the national legislation in the country of data import allows the data recipient researcher to comply with the contractual provisions [11].

However, it is unclear whether this can realistically succeed. Researchers cannot bind themselves to rules contradicting their obligations under domestic law, such as requirements to disclose data to local authorities. Furthermore, in some countries—such as the United States—researchers are often subject to legislation that prevents them from signing the Standard Contractual Clauses. In addition, their potential to enhance data subjects' protection by, for instance, establishing institutional complaint mechanisms has only a limited effect on the actual improvement of the fundamental rights protection of those affected if administrative and judicial remedies in the concerned jurisdiction cannot safeguard the fundamental rights of EU citizens. Hence, even the recent call for special contractual

clauses for the scientific research processing sector [12] falls flat if the regulatory environment does not secure a comparable level of fundamental rights protection for the data subject equivalent to that of the GDPR. In this case, the concerned sector cannot be deemed adequate as other applicable rules (related to compliance with national surveillance mechanisms) preclude researchers from contractually binding themselves to terms that approximate those of the GDPR. As a consequence, research endeavors involving a transfer of personal data from the EU or the European Economic Area to the United States often cannot be implemented in many cases, such as sending deidentified human genetic data to the Imputation Server hosted by the University of Michigan or pooling personal data on a single server as envisaged by the International Alzheimer's Consortium and the US-based Alzheimer's Disease Sequencing Project [13].

Performing an assessment of the law and practice of the jurisdiction of a destination is referred to in the literature as a "mini-adequacy decision" [14]. With each transfer, data exporters are required to consider the ease of access to data by government actors, the possibility for the simplified exercise of rights and effective remedies for breach thereof, and whether the destination of the transfer is governed by the "rule of law." Performing such an analysis in individual cases and monitoring changes in local law and government practices is likely to exceed the capabilities of researchers exporting data, who are obligated to perform the test in the first place. There is still no helpful guidance on how data controllers responsible for defining the purposes and essential means of data processing are supposed to perform a task that the European Commission has failed to tackle on more than one occasion, as evidenced by the annulment of its adequacy decisions by the CJEU. Shifting the burden of an all-encompassing assessment of a third country's legal system to exporters of data might lead to quasi-arbitrary evaluations as well as to divergences in the application of the adequacy criteria from one another [15]. If these evaluative exercises are carried out poorly, it could lead to the erosion of the fundamental rights of the affected EU citizens. This could take a long time to remedy as the CJEU is the only body competent to do so, and it requires more than a year to decide cases [16].

Do Anonymization or Security Measures Offer a Solution?

According to the European Data Protection Board, the group of national supervisory authorities interpreting the GDPR, any anonymization must be completely irreversible. Some further consider that anonymization must be *future-proof* such that anonymization is impervious to new technologies not yet invented. Indeed, the European Data Protection Board presents the deletion of original data and the removal of characteristics as ideal technical measures of anonymization [17,18].

An exigent threshold for what constitutes anonymized data is a barrier to international genomic research. In health-related genomic data processing, this approach poses acute difficulties as the interrogation of disease etiology and other determinants of health requires personal data. Deleting or stripping data sets of certain variables in the name of anonymization is then directly opposed to the very reason for which processing is undertaken. Examples include cancer imaging data—the anonymization of head and neck images constitutes a serious challenge to the preservation of essential scientific data; as such, modification of the data can diminish their scientific quality and utility [19]. In addition, removing metadata of a specific format (eg, Digital Imaging and Communications in Medicine) [20] from cancer imaging data sets that may constitute indirect identifiers (eg, the manufacturer's serial number) would imply the loss of traceability of the patient. A loss of traceability can have particularly severe consequences in international clinical trials where the ability to follow patients is essential and the identifiable verification of the study results constitutes a legal duty in many countries; their return to the patients must represent an inherent part of the study concept [21]. Furthermore, data anonymization methodologies can systematically deprive the members of small population groups, including traditionally marginalized groups, from inclusion in scientific data sets. This is the case as the indirect identifiers of small population groups are less common than those of majority groups and, therefore, deidentification methods tend to remove them from data sets more often [22].

As a result, the advantages of anonymization cannot be realized in a research context without drastically reducing the potential of the research. Decreasing the richness of data diminishes their scientific value, further limiting the research questions that the data can address, the applicable research methods, and the relevance of the research findings. In addition, divergent data quality will ultimately reduce interoperability between data sets and may even affect the reproducibility and comparability of research results, destroying their statistical validity [23].

Methods other than data anonymization, such as coding and encryption, cannot necessarily facilitate the international transfer of personal data. These methods do not preclude the application of the GDPR to the data. The European Data Protection Board considers coding and encryption methods to be supplementary measures enhancing data protection compliance efforts rather than anonymization techniques that render data nonpersonal.

It seems that security measures widely used in genomic research, such as pseudonymization, cannot remedy this issue, either. The European Data Protection Board considers encryption methods and coding, such as pseudonymization, to be supplementary measures enhancing data protection rather than measures that render data nonpersonal and, thus, outside of the material scope of the GDPR [24]. Accordingly, GDPR requirements for transfer cannot be fulfilled in most cases if it is not possible to protect encrypted and coded data against large-scale access and monitoring by the third countries' law enforcement agencies without corresponding administrative and judicial remedies.

Consequences of the Current Rules

In summary, it can be stated that all mechanisms offered by the GDPR to secure admissibility of international data transfers can only be applied if the recipient outside the EU or the European Economic Area provides an *essentially equivalent* level of data

protection. The fundamental rights context surrounding data processing in the recipient country will influence the assessment of whether rules defining data processing in a certain sector are essentially equivalent to the GDPR standard. This means that the burden of investigating adequacy in recipient countries without an adequacy decision by the European Commission will ultimately lie with the data exporter. At the same time, data recipient researchers will need to determine whether they can sign the offered GDPR Standard Contractual Clauses and adhere to them or whether there exist contradicting obligations for them based on national laws. These rules place a burden on researchers in an era where compiling large data sets across cohorts and countries is crucial for achievements in genomic science.

What happens currently to a data transfer to a third country without an adequate level of data protection? In the absence of an adequacy decision and in the event that Standard Contractual Clauses or other transfer mechanisms cannot ensure an essentially equivalent level of data protection in the country of the data recipient, the controller must provide additional security measures that effectively prevent external access to the data and, thus, protect the rights of data subjects. The function of additional measures is similar to that of any transfer mechanism: they should *compensate* for the lack of high-level data protection that is essentially equivalent to that in the EU and the European Economic Area. These additional measures include the anonymization of all personal data or privacy-preserving techniques such as encryption and coding, where only the data exporter has the key and which cannot be circumvented by others [24].

Anonymization is not a viable solution to circumvent the application of data protection rules. On the one hand, genomic data are highly identifiable because of further data linkages. In contrast, the benefits of genomic research in a health context can only be achieved regularly if there is at least a stratified possibility of tracing the data back to the affected patients and probands.

Concerning privacy-preserving techniques, a discrepancy between technical measures and the standard of "essential equivalence" emerges. Adequacy is fundamentally a legal standard that includes considerations of data access by authorities and the legal obligation of data importers to comply with access orders. Depending on the legal safeguards and available redress mechanisms, the order and corresponding obligation to comply with it may *in themselves* create a processing context for data importation that is below the standard of the EU. Furthermore, issues such as encryption and the availability of other data to bypass the contextual anonymity offered by pseudonymization (coding) are technical. Overturning technical data security will allow for the application of a processing context that would render protection inadequate. Although researchers will only be able to influence technical measures applied to their data processing, the lawful access by the recipient countries' authorities and its interrelatedness with the technical factor needs to be dealt with.

Emerging Solutions to Lawful Data Access by Third-Country Authorities

Generally, the task of assessing whether the level of data protection in a third country is equivalent in substance to the level under EU law is not a mechanical exercise but must involve sophisticated analysis of the legal order of the third country. The analysis must not only cover all areas of law in terms of legislation and case law but also further extend to administrative practices. That is, a study of the literal text of the law is insufficient. Facts on the ground, such as actions taken by administrative bodies, also matter. Evaluating the conservatory measures that the recipients of international data transfers take against orders or requests for information from law enforcement agencies and surveillance bodies could inform the assessment of the "essential equivalence" of a recipient jurisdiction's legal system and afferent practices.

In Canada, for example, some entities make it clear that they will only comply with a valid court order from law enforcement agencies. When reading the transparency reports of these organizations, it becomes clear that most requests are not legally authorized as law enforcement agencies are essentially "asking" for access but cannot compel it [25]. Even if there is a court order, it might only cover limited data sets independent of the collective access for which a law enforcement agency has asked. Canadian human-participant research norms, which are binding on federally funded research, underscore the obligation of both researchers and their institutions to uphold promises of participant confidentiality, which can require researchers and research institutions to contest court orders for data [26].

Although there is little that can be done about surreptitious surveillance, procedures requiring data recipients to contest authorities' requests and orders for access to personal data should be considered when determining whether an international transfer respects the fundamental rights of data subjects. We believe that the analysis of administrative practices that protect fundamental rights will reveal considerable similarities between EU and non-EU legal systems, which are more telling than the superficial differences arising from the formal comparison of the texts of EU and non-EU data protection legislation.

Countries to which the European Commission has denied GDPR adequacy status or that have had their adequacy status overturned by decisions of the CJEU have taken action to heighten the protection of fundamental rights that is accorded to their citizens. In the United States, data protection rights such as the "right to deletion" are now acknowledged in case law [27]. In addition, some countries that did not offer administrative and judicial data protection remedies to foreign citizens are starting to do so [28]. In Japan, independent administrative oversight has hitherto only been acknowledged in the context of private sector use [29]. Soon, a novel adequacy decision applicable to personal data processing by public sector researchers might be implemented in Japan, extending protection with regard to the public sector use of data. Altogether, significant progress will still be required to raise fundamental rights protection in relation to public authority surveillance to a globally standardized level. Data sharing for scientific research

and health care purposes is a pressing global health concern and a predicate for achieving health equity. The pursuit thereof cannot be made contingent on global consensus on issues of State surveillance. Precluding the international exchange of genomic and health data for political reasons condemns at-risk populations to bear poor health outcomes to place pressure on governments to align on surveillance policy.

Scientific Research: Its Technique and Legal Status

Once within the scope of the application of sector-specific data protection law, it is primarily the context of genomic research that will guide the—often complicated—weighing exercise of competing rights and interests, such as between research freedom and data protection, both of which are fundamental rights capable of being limited. In addition, other legally relevant positions on the side of the data subject, such as their right to health and their right to decide what to do with their data (right to private and family life), may move the metaphorical scale in favor of genomic science in certain contexts. In genomics, affected patients may have significant, real opportunities to benefit from research findings, for instance, the clinical validity of a genetic mutation or when a variant is confirmed through translational scientific research.

It is an outstanding achievement as to how far this weighing has been enhanced on the legislative level in the GDPR, where the emphasis on research freedom is strongly guided by the relevance of scientific endeavors in the public interest and permits the data subjects' data protection interests to be limited while at the same time striving to minimize risks for their privacy [30]. Under the GDPR specific regime for scientific research, the primary role of data security is to mirror the outcome of the trade-off between the main interests of processing intrinsic to scientific freedom and those of privacy in the context of research, with other important interests such as those of the public guiding the trade-off. Obligations for data controllers and processors to implement technical data security measures as defined by the GDPR generally and for scientific data processing specifically (eg, Article 89 safeguards that mandate data minimization and pseudonymization) address the trade-off result between these very interests and rights at the legislative level. Altogether, supplementary measures completing data protection in the genomic data governance context, particularly technical security and administrative protections against enforced data release, and the oversight of future use introduce a cumulative practice of good governance rather than a "silver bullet" consisting of a singular method that alone guarantees data protection, as is implicit in the discussion of secure multiparty computation and other technological measures in recent European guidance [24].

Scientific Research as an Element of Legal Weighing Exercises

Weighing competing rights and interests and translating the result of this balancing exercise into practices, policies, and technical measures enabling secure genomic data exchange will become more complex in the future. Historically, the analysis has required the bilateral consideration of the individual interest in the protection of patients and research participants relative to the research freedom of scientists and the broad societal interest in advancing research and delivering a high standard of health care. However, the relevant interests are now becoming multipolar. Namely, the balancing of interests to be performed becomes multipolar. It becomes necessary to consider not only the privacy interests of individuals relative to scientific freedom and the public interest in scientific progress but also the additional complexities of State surveillance directed at individuals and the public.

When the adequacy of a third country's data protection is contested on the basis of law enforcement's potential access to data and the removal of data anonymity, the initial weighing context related to scientific research and the GDPR's privileging of data processing for scientific research purposes shifts considerably.

The essentially bilateral relationship between privacy and scientific freedom until then determined the appropriate security measures to be implemented. However, where there is a prospect of surreptitious State surveillance or of law enforcement access to data, the security measures and other safeguards to be implemented differ. The choice of appropriate security measures and other safeguards must account for the potential for State surveillance or law enforcement access to data, which is a different analysis altogether. In these circumstances, the assumption of contextual anonymity that undergirds the governance of data for scientific purposes might be more easily dissolved than assumed, especially in relation to the identifiability of genomic data and with the technological tools available to law enforcement [31].

Furthermore, it seems misleading to frame data protection obligations within the binary distinction between anonymized data and personal (including pseudonymized) data. As already described, the anonymized-personal binary is not a pertinent distinction in the context of the decision as to whether processing data for scientific research purposes and transferring them to third countries can occur in a manner compliant with the GDPR. It is personal data to which the rules of the GDPR apply, including rules on transfer. The risk of identifiability is thus implied in the data security and transfer mechanisms of the data protection law and the rules of the GDPR.

The issue here is that the main technical measures that would still enable meaningful and beneficial genomic science are contradicted by considerations surrounding the legal and de facto possibility of law enforcement circumventing contextual anonymity. We believe that there needs to be a distinct legal framework enabling scientific research as the currently proposed solutions pose challenges to researchers that they cannot solve on their own (Table 1). This framework should protect, where necessary, scientific research from other interests. The creation of such a framework should be based on a normative decision instead of its facilitation being only dependent on the technical agility of data exporters and importers. The main concept of the proposed legal framework is elaborated on in the following sections. Significant elements of the concept are the further

development of contractual obligations for data importers, creating safe data spaces, and working toward linking these with data infrastructures worldwide.

Table 1. Identified challenges related to international data transfers, solutions currently explored, remaining challenges, and proposed solutions to these challenges.

Challenges	Explored solutions	Remaining challenges	Suggested solutions
Lack of adequacy decision by the European Commission	Creation of clear rules for the adequacy assessment procedure	Blanket adequacy decisions easily disregard sectoral differences in applicable data protection rules	Adopt specific health research sector adequacy assessments to take into account specific trade-offs among rights, appropriate technical measures, and long-standing compliance efforts within the sector, including administrative measures
Data anonymization and privacy-preserving data security measures are promoted as the only solutions to data protection concerns	Nonapplication of data protection rules and instead use substitute measures to meet the adequacy standard	Loss of information content of data for scientific research, anonymity of data is context dependent, and substitute measures for protection can be circumvented	Emphasize the contextual anonymity of data, for example, when the context is not changed during data processing (eg, by allowing data to be visited)
Missing codes of conduct and certification mechanisms	Bottom-up sector-specific concretization of data protection rules and appropriate supplementary security measures	Current solutions that are not relevant to the context of genomic research or are not relevant for international data transfers [32]; fundamental rights issues raised	Link the development of codes of conduct with the sectoral adequacy assessment and the development of certification mechanisms with supplementary technical measures for international data transfers

The Way Forward I: Adapting Contractual Settings

To perform a data transfer from the EU or the European Economic Area to a third jurisdiction that does not meet the "European essential guarantees," it is necessary to apply effective supplementary measures that raise the standard of data protection to that enshrined in the GDPR. These measures can include a combination of technical measures, private law arrangements, and organizational practices.

In the absence of effective supplementary measures, supervisory authorities will bear the obligation of determining whether contractual and organizational best efforts to mitigate the potential for surveillance bodies and law enforcement agencies to make surreptitious use of data should be considered sufficient to enable the international transfer of data [31]. Potentially relevant measures include the integration of contractual language mandating mutual transparency to agreements between data importers and exporters. This might include the obligation to regularly provide specific information about requests received from authorities regarding personal data processed under the relevant contract. If disclosing specific details about such requests is otherwise prohibited, general information could still be provided (eg, warrant canaries) [33].

Moreover, the inclusion of obligations specific to the data importer merits consideration. These could include an obligation to take legal action to challenge an order to disclose personal data until all pathways to do so have been extinguished. Precedents for such measures exist in Canadian research ethics guidance [26]. These recommendations are often paired with suggestions for creating joint liability between the data exporter and the recipient as well as with rules for compensation, such as the inclusion of an obligation for the data importer to indemnify the data subject, regardless of fault, against all damage caused by access to the data subject's data by entities of their State. Issues related to the effective enforcement of such additional clauses remain open. Data exporters can demonstrate and document their intention to act in a legally compliant manner by observing the requirement of the supervisory authorities and contacting the particular data importer to arrange for these changes to be made to the provisions of the contractual clauses. The stepwise escalation of discretionary measures by supervisory authorities may eventually reduce the exposure of data exporters acting in good faith to high penalties. However, it does not per se lead to an improvement in the position of data subjects as the demonstration and documentation of the will to comply with the law by the contractual parties does not necessarily guarantee enforceable rights for data subjects.

Furthermore, the proportionality of such extensions to contractual agreements will often depend on whether the data importer is replaceable in the short and medium term by an importer who may more easily guarantee an adequate level of data protection. However, this assessment standard is a double-edged sword. Although the irreplaceability of a data importer might be a good yardstick for further enabling standards in the economic sector, the irreplaceability of a scientific cooperation partner must always be judged against the backdrop of the high standards to which the exercise of the fundamental right to scientific freedom is linked. The risk remains that the replaceability of different partners in genomic science might quickly be based on a superficial comparison of available technical equipment or external indicators of success. Although these metrics can influence the exercise of scientific freedom, they must not influence the protective value assigned to scientific freedom as a *right of freedom*.

The Way Forward II: Safe Data Spaces for Scientific Research

With the European Data Strategy, the EU aims to create a single space for data that will allow them to flow freely within the EU and across sectors for the benefit of businesses, researchers, and public administrations. One of the core pillars of this strategy is precisely the promotion of "Common European data spaces in strategic sectors and domains of public interest," including the European Health and European Research Data Spaces [23]. The EHDS aims to enable an efficient exchange of and direct access to different health-related data across the EU in compliance with data protection regulations, in particular the GDPR [34]. As for the regulatory subject matter, the design of the rules for data exchange as well as connecting the EHDS with the emerging genomic research infrastructure is of particular relevance. The "1+ Million Genomes" is an initiative of individual EU member states that aims to enable the sharing of at least one million genomes by 2022 [35]. This initiative is particularly important as the EHDS expressly includes genomic data in its scope and should be connected to the "1+ Million Genomes" initiative in this regard.

A series of measures are proposed to foster these spaces, including the deployment of data infrastructures, tools, and computing capacity by way of scaling and interoperating repositories and databases in a federated manner [36].

Concentrating secure data processing in a cloistered data space might alleviate the imperfect regulatory environment applicable to research data processing. However, the long arm of the law enforcement regulations of third countries creates difficulties in securing the fundamental right to data protection throughout the entire life cycle of research data processing. Therefore, additional settings may be needed to help maintain data processing within a safe environment, such as preventing the download of data as a technical safeguard accompanied by the legal safeguard of contractually prohibiting it. A further step toward upholding a safe environment for research data processing is to offer a searchable metadata basis without moving data, deploying data analysis services that allow for the submission of research questions, and completely foregoing access to the research data themselves. Federated data sharing models that could ground such development are being successfully implemented by international research data archives such as the European Genome-Phenome Archive [37] and by consortia such as the European-Canadian Cancer Network [38].

In addition to processing data for primary health care purposes, the establishment of the EHDS for secondary data processing is linked to making electronic health data, health-related data already stored by various data holders, and data whose influence on health is known, such as genomic data, widely available for the purposes of health research to various data users [39]. Such data holders include public and private research institutions [40]. The draft EHDS regulation obliges data holders to make the categories of electronic data listed in the regulation available for secondary use [41]. The term "making available" means making the data available to a so-called Health Data Access Body at its request [42]. In addition, data holders are obliged to provide the Health Data Access Body with a general description of the data sets they store [43].

Developments by the EHDS for Secondary Use of Genomic Data for Scientific Research

EU member states are required to appoint or establish public bodies entitled Health Data Access Bodies [44]. Health Data Access Bodies receive and review data users' requests for access to data that are retained in the EHDS for secondary use, including scientific research use [45]. Prospective data users must submit requests to Health Data Access Bodies, which decide whether to authorize access to the requested data [46]. In administering such requests, Health Data Access Bodies assess a number of factors stipulated in the legislation. Relevant considerations include whether the applicant intends to use the requested data for a purpose that the law authorizes, whether the legislative preconditions to data access have been fulfilled, and whether access to the requested data is necessary for the applicant to fulfill their stated purpose [47]. If an applicant fulfills the preconditions of data access, the Health Data Access Bodies must issue a data permit in favor thereof. The permit explicitly establishes the conditions according to which the data can be used. Such a permit is valid for a maximum of 5 years [48].

The requested data are provided in either pseudonymized or anonymized form. Insofar as it is possible for the recipient to achieve their purposes in reliance on anonymized data, the data will be made available to them in an anonymized form. In all cases, data users are strictly prohibited from reidentifying the data that are provided to them [49,50].

Data access is provided through a secure data processing environment. This secure environment is subject to legislatively established security and interoperability requirements. This environment implements the technical and organizational measures required by the GDPR. For example, data users are prevented, through technical controls, from downloading data that are held in the secure processing environment [51]. The proposed legislation does not consider Health Data Access Bodies to be mere stewards of the data that are made available to data users. Rather, both the data user and the Health Data Access Body share legal responsibility for ensuring the lawful use of the requested data—the law considers them to be "joint controllers" [52].

The European Commission will collaborate with member states to create a central infrastructure that enables data users to access cross-border data through national points of contact. Member states can appoint their coordinating Health Data Access Bodies as their respective national points of contact. These contact points will become the authorized participants in the infrastructure [53,54]. National points of contact in each EU member state will compile and publish a holistic, EU-wide catalog of available data sets. This will assist prospective data users in discovering relevant data sets that are held in other EU member states for the purpose of requesting access thereto [55].

Opening Up the EHDS for International Scientific Collaborations

The European Commission intends to enable third countries and international organizations to integrate their own national points of contact with the EHDS infrastructure. The Commission, together with the representatives of the national points of contact of EU member states, referred to as the "joint controllership group," must perform a compliance assessment before admitting foreign nodes to the overall EHDS network [56]. If the outcome is favorable, the European Commission will adopt an implementing act, which states that the concerned foreign node is compliant with the EHDS regulation and further requirements for the secondary use of data and provides access to data users located in the EU to the electronic health data it has access to on equivalent terms and conditions [57]. Thereupon, the foreign node is admitted to the EHDS infrastructure and joins the national nodes of EU member states.

The proposed EHDS legislation establishes specialized rules applicable to the secondary processing of health data for scientific research purposes. These are compatible with the more general GDPR rules that require data protection interests to be balanced against the research interests pursued. To this end, the GDPR requires the necessity and proportionality of the intended data processing to be assessed and considered relative to the sector-specific objectives thereof.

In admitting non-EU infrastructure nodes to the pan-EU network of national points of contact, the European Commission submits the applicant foreign nodes to the aforementioned assessment procedure. By our reading, this assessment procedure mirrors the "adequacy" review that the European Commission undertakes before establishing that data importers in a third jurisdiction are authorized to receive personal data transfers from the EU without such transfers requiring additional legal compliance measures. Therefore, before foreign nodes are integrated into the EHDS infrastructure, it will be necessary for the applicant nodes from third countries to demonstrate compliance with the overall requirements of the GDPR, the EHDS regulation, and the EU fundamental rights framework to the satisfaction of the joint controllership group. This is contingent on a thorough assessment of the legal rules and practices in the applicant's jurisdiction as regards State surveillance, among other factors.

The proposed regulation creates a relationship of joint controllership between EU national points of contact and their non-EU corollaries. This enables data subjects in the EU to assert legal claims against their own respective national EU points of contact for misuses of data that occur through the fault of non-EU points of contact. This may lead to positive outcomes in facilitating access to legal remedies for EU data subjects. However, EU points of contact could be held liable for the activities of their non-EU partners through no fault of their own, including through the breach of EU fundamental rights that arise because of the surveillance activities of non-EU State actors. This prospective liability could have a chilling effect on the joint controllership group that is responsible for determining whether foreign nodes should be admitted. That is, national EU nodes might hesitate to admit foreign nodes to the larger network if the behavior of the foreign nodes could cause the national EU nodes to be held liable for a breach of the GDPR or the EHDS regulation.

The use of a compliance assessment that mirrors the GDPR adequacy procedure to admit foreign nodes to the network of national points of contact of the EHDS is a curious legal design choice. The EHDS technical platform is anticipated to integrate secure data processing capabilities that preclude data from being externally downloaded or otherwise replicated. Regardless of the legal data protection norms—and surveillance practices—applicable in the country of origin of the contributed data, the technical design of the EHDS should achieve a common, GDPR-compliant standard of data protection guarantees. Therefore, it should be further examined whether the policy choice to require a comprehensive compatibility assessment, akin to a GDPR adequacy determination, before integrating foreign nodes into the EU network is justified at all.

The integration of national health data spaces into a larger international network will require governments and regulators to pioneer novel legislative and nonlegislative measures. In this respect, the European Commission holds a rarefied role as both lawmaker and pioneer of critical international infrastructure [12]. The European Commission has previously been criticized for not considering existing measures that are used to balance the risks and benefits of scientific research in performing adequacy assessments directed at the health sector. Perceived ambiguities arise in the guidelines of the European Commission as to the criteria that must be used to determine whether the norms of third countries should benefit from a favorable adequacy decision. This creates legal uncertainties regarding the functioning of the adequacy regime, which is the central mechanism that enables third countries to benefit from unencumbered transfers of data from the EU [58]. It remains to be seen whether the European Commission will implement transparent, comprehensible, and internally consistent methodologies in deciding on the accession of third countries' infrastructures to the EHDS.

The GDPR continues to apply to data processing in the EHDS. Therefore, it remains open to member states to implement supplementary conditions that are applicable to data processing and international data transfer in the EHDS. The potential for member states to do so is bounded by the limits established in the GDPR. Nonetheless, this could detract from the harmonizing prospects of the EHDS in enabling distinct member states to apply their own divergent national norms to their respective nodes of the infrastructure. For example, member states can use domestic law to expressly establish limits to the transfer of specific categories of personal data to a third country or international organization for important reasons of public interest. Such limits may be imposed so long as the concerned country or international organization does not already benefit from a GDPR adequacy decision [59].

In summary, the legislation creating the EHDS reprises numerous restrictive and limitative elements of the GDPR that will continue to impede the potential to make plentiful use of data for genomic research supporting health research and care.

In this respect, the EHDS will likely replicate, not resolve, the problems that the GDPR has created for international biomedical and genomic data exchange (Textbox 1). However, a foundational pillar thereof has been unduly neglected: the seamless integration of international data spaces into the EU infrastructure and convenient access to the data in the EHDS by researchers worldwide.

Textbox 1. Improving the European Health Data Space (EHDS).

Ways of improving the EHDS

- Interpret its data sharing rules against the backdrop of necessity and proportionality of data processing for scientific research purposes
- Create detailed rules for international joint controllers assigning clear obligations to best secure the data protection rights of patients and participants
- Relieve the burden of the main rule of data anonymization for scientific data processing not to affect the quality and usefulness of research results
- Improve security and organization through further measures such as implementing data analysis services
- Reduce member states' individual rules for data sharing through sectoral harmonization by means of certification mechanisms and codes of conduct
- Acknowledge making data available through the EHDS as a data processing step that does not constitute an international data transfer
- Acknowledge its security and organization as data sharing that is adequate for the genomic sector
- Foster public interest in genomic science through participation, information, and transparency

Other as-yet Unused Policy Instruments to Support International Data Sharing

Having addressed how lacunae in the present draft of the EHDS legislation could inhibit equitable collaboration in international research, we now consider prospective alternatives to the current design.

The mandate to create searchable, nonpersonal data catalogs is a positive development that will help make data findable for scientific research across regions and countries. However, the EHDS legislation goes on to establish that primary data access—rather than simple data discovery—will also require the accessed data to be anonymized if identifiable data are not strictly necessary for the intended purposes. Performing scientific research using anonymized data inhibits the prospect of gaining knowledge through the analysis thereof and inhibits generalizable conclusions from being derived therefrom that can be applied to patient care.

Considering that the EHDS intends to restrict data processing to a cloistered technical infrastructure that does not enable users to download or otherwise duplicate the concerned data, the additional presumption in favor of data anonymization appears overzealous. It pursues duplicative privacy controls at little anticipated gain for data subjects while deprecating the anticipated discoveries that scientific research communities can derive through the analysis of data. At the same time, it is not comprehensible why no distinction is made within the EHDS between the assessment of the data protection standard for international scientific collaborations based on the processing of anonymous data that do not fall within the scope of data protection laws and deidentified or pseudonymized data that do.

It is recommended that the access of researchers in non-EU countries to the EHDS not be treated as an international data transfer for the purposes of the GDPR. The GDPR applies additional rules to international transfers of personal data that are directed at non-EU jurisdictions (or, rather, jurisdictions outside the European Economic Area). These rules are implemented to ensure that the standard of data protection—and the fundamental rights guarantees—that is ensured to EU data subjects is not compromised through the transfer of such data to different countries that incorporate different—and potentially lower—thresholds of data protection to their own national norms. As access to data in the EHDS is performed on EU infrastructure according to technical specifications determined by EU policy makers, it is appropriate to avoid treating such data processing activities as outbound data transfers from the EU. Indeed, there is no prospect for such data processing activities to inhibit the data protection guarantees provided to EU data subjects as both EU and non-EU access to data hosted in the EHDS take place according to the same conditions.

This determination is consistent with the CJEU's jurisprudence, the highest court of the EU. As early as 2003, the CJEU stated that it could not be presumed that the expression "transfer [of data] to a third country" intended to include the loading of data onto an internet page even if those data were thereby made accessible to persons in third countries. If this provision were interpreted to mean that there is "transfer [of data] to a third country" every time that personal data are loaded onto an internet page, that transfer would necessarily be a transfer to all the third countries where there are the technical means needed to access the internet. This special regime would thus necessarily become a regime of general application with regard to operations on the internet [60].

The European Data Protection Board has since issued guidance that seems to contradict the foregoing case law. The Board states that an international data transfer includes not only the outright transmission of data to third parties but also acts that *make the data available* to different actors or entities in third countries regardless of whether such importers are subject to the GDPR with respect to the concerned processing activities [61].

It is uncontentious that accessing genomic data stored on an EU platform via the internet is considered a data processing operation. However, further clarification is required as to

whether mere non-EU access to EU-hosted data constitutes an international data transfer. That is, the GDPR international data transfer rules are intended to be drawn on in cases where the application of non-EU data protection rules and government practices to EU-derived data has the potential to erode the privacy and data protection guarantees to which data subjects in the EU are otherwise entitled. If the EHDS data platform creates a safe data space through technical measures and data visitation requirements that ensure the continued application of EU data protection standards, it stands to reason that data processing performed on such a secure platform would not trigger the application of GDPR data transfer rules.

Regulated entities can adopt specialized tools to tailor the application and interpretation of the GDPR to a particular economic sector or sphere of activities. These include codes of conduct and certification mechanisms, among other similar tools. Implementing these mechanisms in the context of genomic research could help facilitate the outbound transfer of such data from the EU. Indeed, this is the case as the GDPR recognizes compliance with codes of conduct and certification mechanisms that the European Commission has approved as methods of ensuring the lawful outbound transfer of data from the EU to non-EU jurisdictions even in the absence of an adequacy decision in favor of the country of destination [62]. However, as with all transfer instruments intended to compensate for a lack of adequacy, the rules of such codes of conduct or certification mechanisms must be observed through binding and enforceable commitments on the part of the data recipient in the third country. These must bind the data recipients to the conditions established in the code of conduct or certification mechanism and must further guarantee respect for the fundamental rights of EU data subjects. These mechanisms bear the same limitations as other GDPR transfer mechanisms regarding the fundamental rights of EU data subjects. That is, none can overcome State surveillance practices and discrepancies in local law that would enable State actors to access the data of EU data subjects despite binding and enforceable commitments not to share such data entered into by the data recipient.

Therefore, both codes of conduct and certification mechanisms can suffer from the same imperfect fundamental rights environment as any other GDPR transfer mechanism. Despite these limitations, the aforementioned transfer mechanisms are always created in a sector-specific manner that helps specify the application of data protection rules to the particularities of the concerned data processing activities. This helps identify the technical data protection measures that are relevant to the processing activities of the concerned economic sector and balance data protection interests against other competing interests in a context-sensitive and sector-relevant manner.

Conclusions

Providing sector-specific, purpose-related rules through codes of conduct and clarifying the boundaries of the term “transfer” in data protection law will contribute to nuanced international data sharing rules. Indeed, in carefully narrowing the ambit of international data transfers to those uses of data that pose a prospective risk to the fundamental rights of EU data subjects, EU regulators will incentivize the design of legal and technical enclaves enabling non-EU data users to process EU data in a manner that benefits EU and non-EU communities without engendering correlative risks to individual privacy. However, ultimately, both the international community and individual countries are called upon to collaborate in raising local standards of data protection to provide minimum guarantees against State surveillance that are compatible with human and fundamental rights. At the same time, it is neither fair nor necessary to inhibit data use that enables genomic research because of incompatibilities in national legal systems protecting data subjects from surveillance and incompatibilities that arise outside the context of scientific research.

Determining the appropriate boundaries between the privacy rights of research participants and the countervailing exceptional right for State actors to access personal data that have been processed for scientific research purposes to further the interest of law enforcement raises contentious issues of public policy. A delicate balance between the public interest in scientific research and the countervailing interest in law enforcement must be achieved. Interestingly, we already see this in the context of the United States, with Certificates of Confidentiality available to protect participants from forced data disclosure by law enforcement officials [63].

Parallel progress must ideally be pursued in both of the foregoing policy arenas. That is, paths to the secure exchange of biomedical data for research purposes must be negotiated absent global consensus as to the appropriate balance between security or law enforcement interests on the one hand and data protection or privacy on the other. However, at the same time, further international dialogue must be pursued to foster an agreement on a shared minimum standard of data protection and privacy rights for individuals worldwide. To achieve this objective, it would be possible for EU regulators to issue an adequacy decision in favor of the research sector of a third country. The GDPR provides the possibility of proffering an adequacy decision in favor of only one or more specified sectors within a third country. There are good policy reasons for pursuing this path. Indeed, there have been considerable efforts on the part of scientific research communities to ensure the good governance of collaboratively generated scientific research data. International collaboration has contributed to the development of common data stewardship practices, data security standards, and biomedical research ethics rules throughout global biomedical research endeavors. It is up to lawmakers to acknowledge these efforts and bridge the gap by providing the corresponding sectoral protection of data sharing and ensuring that its processing purpose remains for scientific research in the public interest shielded from fundamental rights intrusions.

Such a development would, in the short term, constitute an appropriate recognition by lawmakers of both the positive and negative dimensions of freedom of scientific research. From a negative rights standpoint, this would protect researchers from State incursions on this fundamental right. From a positive rights standpoint, this recognition would impel scientists to pursue the dual objectives of protecting data subjects’ rights and freedoms while also excelling in the production of

state-of-the-art research outputs. In the medium term, the creation of "safe data spaces" can contribute to the efficient pursuit of scientific advancement, creating a favorable regulatory environment that enables contribution to and benefit from existing scientific data resources on the part of scientific communities and the general public in compliance with clearly defined legal preconditions. In the long term, the advent of safe data spaces can significantly contribute to the formation of a novel regulatory sector in the health sciences that directs public and private resources toward judiciously balancing the interests of the main contributors and stakeholders engaged. These stakeholders include patients, research participants, researchers, and physicians. Thus, the legislator would act as a focused enabler. These developments would ultimately foster the development of a sector-specific adequacy standard in the area of health research, the foundation of which is already established in the GDPR. These considerations should serve as the beginning of a robust and global health data governance framework with standardized and binding international rules for scientific health research, including genomic science, developed and implemented in all of our interests as a global community.

Acknowledgments

This project has received funding from the European Union Horizon 2020 research and innovation program under grant agreement 825835 and the European-Canadian Cancer Network (EUCANCan), a federated network of aligned and interoperable infrastructures for the homogeneous analysis, management, and sharing of genomic oncology data for personalized medicine. MB, AB, and BMK received funding support for this work from the Canada Foundation for Innovation Cyberinfrastructure Initiative; the Ontario Research Fund; the British Columbia Knowledge Development Fund; and the Ministère de l'Économie, de la Science et de l'Innovation for *The Cancer Genome Collaboratory* project, as well as from the Canada Research Chair in Law and Medicine. FMG is funded by the Deutsche Forschungsgemeinschaft (German Research Foundation)—NFDI 1/1 "German Human Genome-Phenome Archive." FMG, AB, PN, MRL, and BMK are members of the ethico-legal work package of the EUCANCan consortium. The authors are grateful for the collaborative spirit within EUCANCan and for the open exchange with all members of the consortium.

Authors' Contributions

FMG drafted the manuscript. All authors provided substantial contributions to the manuscript. All authors approved the final and revised manuscript before submission.

Conflicts of Interest

FMG is a member of the European Group on Ethics in Science and New Technologies. The views presented in the manuscript are not necessarily those of the Group. MB, AB, PN, MRL, and BMK declare that they have no competing interests.

References

1. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature 2020 Feb 05;578(7793):82-93 [FREE Full text] [doi: 10.1038/s41586-020-1969-6] [Medline: 32025007]
2. Regev A, Teichmann S, Lander E, Amit I, Benoist C, Birney E, Human Cell Atlas Meeting Participants. The human cell Atlas. Elife 2017 Dec 05;6:7041 [FREE Full text] [doi: 10.7554/eLife.27041] [Medline: 29206104]
3. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, 4.5.2016, p. 1-88. [FREE Full text]
4. Gourd E. GDPR obstructs cancer research data sharing. Lancet Oncol 2021 May;22(5):592. [doi: 10.1016/s1470-2045(21)00207-2]
5. Judgment of the Court (Grand Chamber) of 16 July 2020, Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems, Case C-311/18, ECLI:EU:C:2020:559 ("Schrems II"). 2020 [FREE Full text] [doi: 10.1007/s40319-020-00967-2]
6. Judgment of the Court (Grand Chamber) of 6 October 2015, Maximillian Schrems v Data Protection Commissioner. Case C-362/14, 6 October 2015, ECLI:EU:C:2015:650 ("Schrems I"), recital 73. URL: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:62014CJ0362
7. Working document on Adequacy Referential (wp254rev.01), p. 2, 10. European Commission. 2018 Feb 9. URL: https://ec.europa.eu/newsroom/article29/items/614108/en [accessed 2022-09-24]
8. Opinion of Advocate General Bot delivered on 23 September 2015. Case C-362/14 ("Schrems I"), recital 141. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62014CC0362
9. Commission Implementing Decision (EU) 2019/419 of 23 January 2019 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by Japan under the Act on the Protection of Personal Information. OJ L 76, 19.3.2019. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019D0419

10. Adequacy decisions. European Commission. URL: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en [accessed 2022-09-28]
11. Bradford L, Aboy M, Liddell K. Standard contractual clauses for cross-border transfers of health data after. J Law Biosci 2021;8(1):lsab007 [FREE Full text] [doi: 10.1093/jlb/lsab007] [Medline: 34164131]
12. International sharing of personal health data for research, p. 33. European Academies Science Advisory Council. 2021 Apr 8. URL: https://easac.eu/publications/details/international-sharing-of-personal-health-data-for-research/ [accessed 2022-09-28]
13. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR's frustration of global biomedical research. Science 2020 Oct 02;370(6512):40-42. [doi: 10.1126/science.abd2499] [Medline: 33004505]
14. Kuner C. The Schrems II judgment of the court of justice and the future of data transfer regulation. European Law Blog. 2020 Jul 17. URL: https://europeanlawblog.eu/2020/07/17/the-schrems-ii-judgment-of-the-court-of-justice-and-the-future-of-data-transfer-regulation/ [accessed 2021-12-14]
15. Stoddart J, Chan B, Joly Y. The European Union's adequacy approach to privacy and international data sharing in health research. J Law Med Ethics 2016 Mar 01;44(1):143-155. [doi: 10.1177/1073110516644205] [Medline: 27256130]
16. The year in review, p. 24. Court of Justice of the European Union. 2020. URL: https://curia.europa.eu/panorama/2020/en/index.html [accessed 2021-12-14]
17. Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN. Adopted on 10 April 2014, p. 21, no. 4.1. Article 29 Data Protection Working Party. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [accessed 2022-09-28]
18. EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research. European Data Protection Board. 2021 Feb 2. URL: https://edpb.europa.eu/sites/default/files/files/file1/edpb_replyec_questionnaireresearch_final.pdf [accessed 2022-10-04]
19. Diaz O, Kushibar K, Osuala R, Linardos A, Garrucho L, Igual L, et al. Data preparation for artificial intelligence in medical imaging: a comprehensive guide to open-access platforms and tools. Phys Med 2021 Mar;83:25-37 [FREE Full text] [doi: 10.1016/j.ejmp.2021.02.007] [Medline: 33684723]
20. Bidgood WD, Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. J Am Med Inform Assoc 1997;4(3):199-212 [FREE Full text] [doi: 10.1136/jamia.1997.0040199] [Medline: 9147339]
21. Getz K, Farides-Mitchell J. Assessing the adoption of clinical trial results summary disclosure to patients and the public. Expert Rev Clin Pharmacol 2019 Jul 15;12(7):573-578. [doi: 10.1080/17512433.2019.1615441] [Medline: 31050914]
22. Wilkinson K, Green C, Nowicki D, Von Schindler C. Less than five is less than ideal: replacing the "less than 5 cell size" rule with a risk-based data disclosure protocol in a public health setting. Can J Public Health 2020 Oct 11;111(5):761-765 [FREE Full text] [doi: 10.17269/s41997-020-00303-8] [Medline: 32162281]
23. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European Strategy for Data, p. 9. 2020 Feb 19. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066
24. Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data. Adopted on 18 June 2021, No. 80 et seq. European Data Protection Board. URL: https://edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-012020-measures-supplement-transfer_en [accessed 2022-09-28]
25. Transparency report. TekSavvy. URL: https://www.teksavvy.com/policies/legal-stuff/transparency-report/ [accessed 2021-12-14]
26. Tri-council Policy Statement Ethical Conduct for Research Involving Humans, Article 5.1. Ottawa: Interagency Secretariat on Research Ethics; Dec 2018.
27. Judgment of the United States Court of Appeal of 17 June 2014. United States v Ganias, No 12–240 (US). The United States Department of Justice. URL: https://www.justice.gov/usao-ct/file/765036/download
28. European Commission and United States Joint Statement on Trans-Atlantic data privacy framework. 2022 Mar 25. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2087 [accessed 2022-06-04]
29. Gehrke N. Alignment of personal information protection laws in Japan. Tokyo FinTech. 2020 Mar 14. URL: https://medium.com/tokyo-fintech/alignment-of-personal-information-protection-laws-in-japan-3d80e93d331c [accessed 2021-05-31]
30. Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Hum Genet 2018 Aug 17;137(8):657-664 [FREE Full text] [doi: 10.1007/s00439-018-1924-x] [Medline: 30120573]
31. Orientierungshilfe: Was jetzt in Sachen internationaler Datentransfer? Der Landesbeauftragte fur Datenschutz und Informationsfreiheit Baden-Wurttemberg. 2021 Sep. URL: https://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2021/10/OH-int-Datentransfer.pdf [accessed 2022-09-28]
32. Code of conduct regulating the processing of personal data in clinical trials and other clinical research and pharmacovigilance activities. Farma Industria. URL: https://www.aepd.es/es/documento/farmaindustria-code-conduct-regulating-processing-personal-clinical-en.pdf [accessed 2022-07-14]
33. Wexler R. Warrant canaries and disclosure by design: the real threat to national security letter gag orders. Yale Law Journal Forum. 2014 Dec 19. URL: https://www.yalelawjournal.org/pdf/WexlerPDF_xpc2ib9k.pdf [accessed 2022-09-28]

34. European Commission. European health data space. URL: https://ec.europa.eu/health/ehealth/dataspace_en [accessed 2022-09-28]
35. European '1+ Million Genomes' initiative. European Commission. URL: https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes [accessed 2021-12-14]
36. European Health Data Space, Presentation 18th eHealth Network, 3.1. DG Santé. 2020 Nov 12. URL: https://health.ec.europa.eu/latest-updates/summary-report-18th-meeting-ehealth-network-12-13-november-2020-2021-02-25_en [accessed 2022-09-14]
37. What is in the EGA? The European Genome-phenome Archive. URL: https://ega-archive.org/ [accessed 2021-12-14]
38. European-Canadian Cancer network. EUCANCan. URL: https://eucancan.com/ [accessed 2021-12-14]
39. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, recital 38 and 39. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
40. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, recital 38 and 40. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
41. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 33. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
42. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 41 (1), (4), recital 40. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
43. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 41 (2). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
44. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 36 (1). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
45. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 37 (1) (a), recital 42. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
46. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 45 (1). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
47. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 46 (1). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
48. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 46 (6), (9). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
49. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 44 (2). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
50. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 44 (3). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
51. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 50 (2). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
52. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 51. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
53. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 52 (2). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
54. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final Art. 52 (1). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
55. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 57. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]

56. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 52 (5), (14) and Art. 66. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
57. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 52 (13) (c), (d). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
58. Charter of Fundamental Rights of the European Union. OJ C 326, 26.10.2012, p. 391–407, Art. 41. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT [accessed 2022-09-28]
59. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final, Art. 63. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197 [accessed 2022-09-28]
60. Judgment of the Court of 6 November 2003, Criminal proceedings against Bodil Lindqvist. Case C-101/01, 6 November 2003, ECLI:EU:C:2003:596 ("Lindqvist"), recital 53 et seq. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62001CJ0101 [accessed 2022-09-28]
61. Guidelines 05/2021 on the Interplay between the application of Article 3 and the provisions on international transfers as per Chapter V of the GDPR. European Data Protection Board. 2021 Nov 18. URL: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2021/guidelines-052021-interplay-between-application_en [accessed 2022-09-28]
62. Phillips M, Molnár-Gábor F, Korbel JO, Thorogood A, Joly Y, Chalmers D, et al. Genomics: data sharing needs an international code of conduct. Nature 2020 Feb 05;578(7793):31-33. [doi: 10.1038/d41586-020-00082-9] [Medline: 32025008]
63. Wolf LE, Beskow LM. New and improved? 21 century cures act revisions to certificates of confidentiality. Am J Law Med 2018 May 06;44(2-3):343-358 [FREE Full text] [doi: 10.1177/0098858818789431] [Medline: 30106660]

Abbreviations

CJEU: Court of Justice of the European Union
EHDS: European Health Data Space
EU: European Union
GDPR: General Data Protection Regulation

Edited by T Leung; submitted 11.02.22; peer-reviewed by L Nweke, S Thiebes, Q Zou; comments to author 03.05.22; revised version received 08.06.22; accepted 19.08.22; published 19.10.22.

Please cite as:
Molnár-Gábor F, Beauvais MJS, Bernier A, Jimenez MPN, Recuero M, Knoppers BM
Bridging the European Data Sharing Divide in Genomic Science
J Med Internet Res 2022;24(10):e37236
URL: https://www.jmir.org/2022/10/e37236
doi:10.2196/37236
PMID:36260387

Viewpoint

Online Symptom Checkers: Recommendations for a Vignette-Based Clinical Evaluation Standard

Annabelle Painter[1], BMBCh, MA; Benedict Hayhoe[1], MD; Eva Riboli-Sasco[1], MA; Austen El-Osta[1], MSc, MPA, PhD

Self-Care Academic Research Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Annabelle Painter, BMBCh, MA
Self-Care Academic Research Unit
Department of Primary Care and Public Health
Imperial College London
Reynolds Building
St Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 207 5943 368
Email: a.painter22@imperial.ac.uk

Abstract

The use of patient-facing online symptom checkers (OSCs) has expanded in recent years, but their accuracy, safety, and impact on patient behaviors and health care systems remain unclear. The lack of a standardized process of clinical evaluation has resulted in significant variation in approaches to OSC validation and evaluation. The aim of this paper is to characterize a set of congruent requirements for a standardized vignette-based clinical evaluation process of OSCs. Discrepancies in the findings of comparative studies to date suggest that different steps in OSC evaluation methodology can significantly influence outcomes. A standardized process with a clear specification for vignette-based clinical evaluation is urgently needed to guide developers and facilitate the objective comparison of OSCs. We propose 15 recommendation requirements for an OSC evaluation standard. A third-party evaluation process and protocols for prospective real-world evidence studies should also be prioritized to quality assure OSC assessment.

(J Med Internet Res 2022;24(10):e37408) doi:10.2196/37408

KEYWORDS

online symptom checkers; clinical evaluation; validation; assessment; standards; third-party assessment; quality assurance

Introduction

The last decade has seen a proliferation of online symptom checkers (OSCs). The pervasiveness of smartphones, tablets, and personal computers has increased the availability of these free and accessible decision support tools that offer on-demand symptom assessment at scale [1]. Although many OSCs products are developed by commercial companies as direct-to-consumer products, several products have been deployed within national health care systems including 'National Health Service (NHS) 111' online and Babylon 'Ask A&E' in the United Kingdom and 'healthdirect' in Australia. These patient-facing general purpose symptom checkers are intended for members of the public to use at home as a decision aid to help inform them about the potential cause of their symptoms and where to seek care.

Despite widespread use of OSCs, there are various concerns about their clinical safety and accuracy [1-7]. A key factor contributing to this uncertainty stems from a lack of consensus regarding an objective methodology or an agreed standard for OSC evaluation.

Although OSCs must comply with Medical Device Regulations [8] and are encouraged to align with evidence standards [9], governance structures for digital health technologies in the United Kingdom and European Union do not stipulate any specific clinical evaluation method or protocol for OSCs [8,9]. In 2020, the Care Quality Commission conducted the first regulatory sandbox focused on digital triage tools, highlighting the following:

> *[D]igital triage tools are not fully clinically validated or tested by product regulators and notified bodies.*

> *We have learned that there is great variation in their clinical performance. [10]*

Most OSCs are registered as Class I medical devices in the European Union and the United Kingdom [3]. Class I status involves self-certification by developers and does not require assessment by a notified body [8]. In the United Kingdom, part II of the UK Medical Device Regulations 2002 requires that Class I products must provide evidence of clinical evaluation and that "the data needs [*sic*] to adequately demonstrate that the product fulfils its intended purpose" [8]. However, neither the Medical Device Regulations nor the Medicines and Healthcare products Regulatory Agency provide detailed guidance on how this should be carried out for OSCs specifically, and they do not stipulate a requirement for objective third-party assessment. This has meant that developers can create their own internal methods for clinically evaluating their products without the need for an objective or impartial assessment to be undertaken.

The National Institute of Health and Care excellence (NICE) in the United Kingdom has published and recently updated a set of evidence standards for digital health technologies [9]. This guidance categorizes digital health technologies into various tiers and suggests appropriate evidence requirements for each. However, this guidance does not include any specific approaches for OSC clinical evaluation methodology.

The lack of consensus on OSC clinical evaluation methodology may also account for conflicting results reported in comparative research studies of OSCs [1,2,4-6,11,12]. Despite appearing to share a similar evaluation approach (eg, using clinical case vignettes to compare OSCs to a 'gold standard' set by clinicians), there is notable variation in the methods used at various steps of the evaluation processes in these studies. These include differences in the number, type, and content of vignettes used; who and how many people input the vignettes into OSCs; how the gold standard diagnostic and triage solutions are chosen; how results are benchmarked against the gold standard; as well as the number and specification of performance metrics used. Inconsistencies of study findings may be further compounded by the low quality of most comparative studies published to date, which are largely observational studies, usually published as grey literature and often by OSC developers themselves, introducing a significant risk of bias [7,13]. As a result, the findings of most vignette-based OSC studies are difficult to reproduce independently, and this applies especially to those studies published by OSC developers.

The need for more robust clinical evaluation guidelines for OSCs has been highlighted in existing literature [7,14]. Suggestions include applying extant evaluation frameworks currently used in mobile health and health informatics to OSCs [14]. Future recommendations should ideally build on these suggestions to inform the development of a standard for vignette-based OSC clinical evaluation methodology. The aim of this paper is to characterize a set of congruent requirements for a standardized OSC vignette-based clinical evaluation process.

The recommendations in this paper were developed through evaluation of primary literature alongside informal discussions with OSC developers involved in clinical evaluation and researchers who have undertaken comparative OSC studies.

Recommended Requirements for an OSC Clinical Evaluation Standard

Robust clinical evaluation guidelines are required to align the processes used by both developers and evaluators of patient-facing general purpose OSCs. The development of a congruent and evidence-based guideline is needed to help provide assurance that OSCs are fit for purpose, promote patient safety, and can help facilitate objective and reliable product comparison and benchmarking.

The variability in results from comparative studies highlights the vulnerability of current vignette-based OSC evaluation approaches. Therefore, any standard for vignette-based clinical evaluation of OSCs will require careful consideration to ensure an objective and robust process is specified, including guidance on how these processes should be implemented and reported in a way that is open and transparent.

Table 1 summarizes 15 key requirements across 7 categories within vignette-based clinical evaluation methodology that could benefit from standardization. These recommendations are intended to guide the creation of a shared standard to be followed by both developers and researchers of OSCs when undertaking a clinical evaluation process using vignettes. Although individual vignette data sets and methodological details may vary according to a given OSC use case, this variation would be limited by the parameters set out in this proposed standard. These recommendations do not represent a standardized third-party evaluation protocol but could be followed in the design of such a process. Third-party benchmarking is discussed further in the 'Recommendations for future work' section.

Table 1. Summary of recommended requirements for a vignette-based online symptom checker (OSC) clinical evaluation standard.

Category	Requirements
Vignettes	• Illustrate the method for determining the minimum number of vignettes required for OSC evaluation. • Specify the minimum information to be included in each vignette. • Provide guidance on determining the conditions, symptoms, or spread of cases to be included and assurance that this appropriately represents the target user population. • Specify vignette origin requirements (eg, simulated vs real-world cases).
Clinician assessment	• Specify the appropriate clinicians (including role, speciality, and seniority) to be used in the assignment of 'gold standard' labels. • Illustrate the method used for compilation of a gold standard (eg, averaged single blinded assessment or consensus discussion).
Triage	• Specify standardized triage categories, including setting and time periods. • Provide guidance on the use of triage gold standards as a range for both urgency and setting.
Differential diagnosis	• Illustrate the method for comparing OSC differential diagnosis list to gold standard dispositions.
Accuracy	• Illustrate the accuracy and safety score calculation method, accounting for outcomes that fall both below and above the gold standard. • Specify the minimum accepted accuracy and safety scores.
Safety netting	• Specify how safety netting contributes to product safety scores.
Inputters	• Specify the minimum interrater reliability scores for inputters of vignettes. • Illustrate the method for determining the number of tests and inputters required for each vignette. • Specify the appropriate characteristics for vignette inputters (eg, medical education level and affiliation with developers).

Vignettes

OSCs are most often validated using a set of clinical case 'vignettes' [4-6,11,12]. Each vignette represents a possible clinical scenario or 'case' and contains information such as key patient demographics, relevant medical history, and symptoms. This method has also been used to assess the reliability of clinician-facing diagnostic decision support tools [15,16] and diagnoses made by clinicians [17,18].

The number of vignettes that should be used during clinical evaluation of an OSC is not defined. There is a lack of guidance about the spread of diseases or presentations that should be included in any given vignette set. There is no guidance on how representative the vignettes should be of the target user population in terms of disease incidence or prevalence and demographics, such as gender, age, and ethnicity, risking an increase in existing inequalities through a lack of inclusion [19]. The minimum level or amount of information that each individual vignette should contain (ie, user demographics, comorbidities, and current medications) is also undefined.

The use of vignettes for clinical evaluation has limitations. Clinicians assessing vignettes are restricted to the information provided without the opportunity to ask additional questions, examine the patient, or assess nonverbal cues. Meanwhile, when an OSC is being tested using a case vignette, the inputter may be forced to make assumptions when answering questions about aspects that are not illustrated in the finite vignette script [12].

Most published OSC comparative studies use disease-based vignettes authored by clinicians [4,5,11,12,16]. These vignettes have additional constraints as clinical authors are likely to describe symptoms differently to patients and vice versa. These imagined vignettes are also subject to bias from authors' clinical experience and education and may result in 'textbook' presentations of diseases rather than realistic cases. The potential for bias is further compounded by the fact that vignettes used in the clinical evaluation of an OSC by developers can be written by clinicians employed by developers themselves. Vignettes coproduced with direct patient input and based on real-world patient-reported symptoms that are not created by OSC developers may be preferable for use in OSC validation studies.

Given the limitations of a vignette-based approach, any OSC clinical evaluation standard involving vignettes should specify the following: (1) the number of case vignettes that must be used to test an OSC; (2) the minimum information to be contained in each vignette; (3) the conditions, symptoms, or spread of cases, that must be included in this data set, including representability to the target user population demographics; and (4) the provenance and creation process of the vignettes (eg, whether they are simulated or real-world cases).

Clinical Assessment

Gold Standard

A 'gold standard' label is a term used to refer to an 'ideal' set of outcomes to which an OSC is compared during a clinical evaluation process. In clinical practice, there is no 'ideal' way of triaging patients and no 'perfect' differential diagnosis. However, gold standard triage and diagnostic labels are required to obtain a quantitative assessment of OSC performance. The 'gold standard' vignette labels used in these assessments are generally assigned by practicing clinicians [4,5,12,20].

The type of clinical professional and the speciality and level of seniority of the clinicians used to generate these labels are all likely to have an impact on the gold standard that is generated. Published studies comparing OSCs to date have used different types and number of clinicians to develop their gold standard labels. Several studies used groups of general practitioners (GPs) [11,12], whereas others used a range of clinical professionals including GPs, paramedics, pharmacists, emergency medicine consultants, and triage nurses [1,4,20]. These differences may have contributed to the varying outcomes of these studies and highlights the need for a consensus.

Gold standard labels are vulnerable to significant interclinician variability even among clinicians from the same field of specialization [12,17,18]. Therefore, the method in which the labels from different clinicians are collated can impact the gold standard. There is a notable disparity in the collation methods used in published studies. Whereas some studies collate assessments by using the majority outcome or the most severe outcome [11,20], other approaches center around asking clinicians to discuss cases together to reach a consensus decision either in a single session or following a series of 'roundtables' [1,12].

Variability in the clinician type and number as well as the methods used to assign and compile gold standards is likely to have contributed to the inconsistency in the results of published OSC evaluation studies. A vignette-based OSC clinical evaluation standard should specify the appropriate cadre of clinicians (including role, speciality, and seniority) and the approach they use to assign gold standard labels and illustrate the method for compilation of a gold standard (eg, averaged single blinded assessment or consensus discussion).

Triage

OSCs may provide users with a triage or priority recommendation advising at what setting and with what degree of urgency to seek help. Urgency refers to how soon a person should be assessed by a health care professional (eg, seek care immediately, within 48 hours, or within 3 weeks), whereas setting refers to the specific area of health care most appropriate for this assessment (eg, emergency department, GP, or pharmacy). Both are important factors when benchmarking a triage recommendation by an OSC to a gold standard.

A major challenge with producing a consensus in the clinical evaluation process for OSCs or in comparing the performance of different OSCs is that different OSCs use different urgency and setting categories [12]. Whereas some OSCs may have urgency categories with a time horizon of 'within 1 hour,' 'within 1 day,' and 'within 1 week,' others may use 'within 6 hours,' within '48 hours,' and 'within 2 weeks.' The same issue applies to the speciality or service setting (eg, some OSCs may suggest pharmacist, dentist, and physiotherapist, whereas others may suggest self-care, GP, and ED). This has led to attempts to map the outcomes from different OSCs to variable reference category sets in comparative studies [5,6,11,12]. Guidance is required to standardize the method of comparing outcomes from different OSCs or to specify the use of a standardized triage category set for both service settings and time horizons. Health systems vary considerably in terms of access to health assessment and advice; therefore, triage recommendations that are appropriate in one country or setting may be unrealistic or unachievable for users in other countries or settings. This will need to be considered in the formation of an evaluation standard.

The use of triage ranges should also be considered when benchmarking OSC performance. Several comparative studies assessed OSCs based on whether they exactly matched a gold standard triage category [1,5,6,12]; however, OSC triage outcomes that are slightly outside of the gold standard may still be clinically appropriate and safe [11]. An example case is as follows: a case vignette describes a patient with ear pain. The gold standard triage solution has been set to 'see GP within 3 weeks.' When tested, the OSC triage recommendation was 'see a pharmacist within 1 week.'

In the example described, this OSC triage recommendation does not exactly match the correct gold standard triage solution; however, it may still be considered safe and would likely result in the appropriate use of health resources. It may therefore be more appropriate for a clinical evaluation standard to outline an approach that encourages the use of a gold standard range for triage solutions of both urgency and setting rather than a singular outcome.

Differential Diagnosis

Benchmarking OSC differential diagnoses to a set of gold standard diagnoses presents unique challenges. The method employed in most published evaluation studies involves using vignettes that are written to represent specific diseases. OSCs are then assessed to see if they suggest this disease as the most likely diagnosis or as part of a differential diagnosis list [1,4,5,11,12].

One significant limitation with this approach is that many OSCs suggest several possible diagnoses, and it is important that each proposed diagnosis is congruent with the case vignette. Secondly, this method is limited when delineating rare conditions from a vignette. This is largely because the symptoms of rare conditions are often also shared with much more common conditions, implying that the 'ideal' outcome for a vignette for a rare disease would not necessarily place the rare disease as the top differential. Ideally, all the OSC differentials should be included in a comparison to gold standard differential diagnoses solutions rather than simply matching a specific disease label.

Safety and Accuracy Thresholds

OSCs are unlikely to always match a gold standard solution exactly. Accepted safety and accuracy thresholds when compared to a gold standard solution, and how such standards should be calculated, will need to be carefully considered in the development of a shared clinical evaluation standard.

In the absence of an agreed standard, developers can set their own safety and accuracy thresholds, which could risk unsafe products being released and causing patient harm. On the other hand, due to concerns about patient safety and product liability, there is also a tendency for OSCs to be risk averse. Studies have

demonstrated that OSCs often advise contact with health care services for conditions that can be self-managed and thus 'overtriage' patients, which could result in an increased burden on health care systems [1,2,4,13,21-23]. Overtriage may also cause unintended harm to patients through heightened anxiety as well as unnecessary investigations and treatments. As such, it is important that the frequency of OSC outcomes that exceed an agreed gold standard triage or diagnosis severity is considered alongside the frequency of outcomes that fall below it. This is supported by the Care Quality Commission regulatory sandbox on digital health triage tools, which suggested that

> *assessments should be based on where people have been wrongly escalated resulting in undue anxiety, as well as where tools have failed to address people's ill health. [10]*

Safety Netting

Some OSCs offer safety netting advice to users in addition to triage and differential diagnosis outcomes. Safety netting includes advice about possible future symptoms that may suggest deterioration or warrant a more urgent health care review. For example, an OSC might suggest that a patient books a routine appointment with their GP within 3 weeks, while also advising that if certain symptoms develop or worsen, they should see a GP sooner or attend ED.

The presence and quality of safety netting is often overlooked in OSC comparative studies, but it is an essential part of traditional doctor-patient consultations and considered during assessments of medical negligence [24]. A consensus clinical evaluation guideline should specify how safety netting should be incorporated into safety and efficacy ratings.

Inputters

Comparative studies showed that inputters can get different consultation outcomes when testing the same vignettes demonstrated by high levels of interrater variability [12,25].

This may be due to variations in how inputters answer OSC questions. For example, one person's interpretation of 'fever' or 'severe pain' may vary from another, causing them to answer questions differently. Research has also shown that inputters may also enter symptoms in a different order, and some may enter an incomplete list of symptoms, both of which can result in completely different OSC outcomes [26].

Comparative studies published to date have used variable numbers of inputters; some have used a single inputter [4-6], while others have used multiple [11,12]. The inputters have also varied in terms of medical literacy, with some studies using qualified medical professionals as inputters [11] and others using nonmedically qualified individuals [4-6,12]. These differences may cause significant variations in vignette interpretation and OSC outcomes. Multiple nonmedically qualified inputters may best represent real-world OSC users.

Given this variation in outcomes depending on individual inputters, a clinical evaluation standard should specify the minimum scores for interrater reliability. This could be combined with stipulating how many independent inputters should be required to test each vignette during evaluation with an average taken of the various obtained outcomes [12]. A defined order for symptom entry and a process evaluating the wording of OSC questions for clarity and ease of interpretation could also be considered.

Recommendations for Future Work

Third Party Benchmarking

In addition to shared clinical evaluation guidelines, an important next step in improving confidence in the safety and accuracy of OSCs would be the development of an objective third-party benchmarking process for OSCs. This has been recommended by the Care Quality Commission sandbox, stating the following:

> *NHSX and NHS England should work with NICE NHS Digital to develop and publish the results of a fair test of clinical performance. [10]*

The results would ideally involve the curation of a set of evaluation vignettes described as a "national dataset of real patient histories, which is not shared with suppliers" [10].

Independent case vignette repositories have also been suggested by authors of comparative studies [6].

Two United Nations agencies—the World Health Organization and the International Telecommunication Union—established a Focus Group on Artificial Intelligence for Health (FG-AI4H) in July 2018. FG-AI4H is developing a benchmarking process for health artificial intelligence models that can act as an international, independent, standard evaluation framework. It has a topic group focused on artificial intelligence–based symptom checkers with participation from numerous OSC developers, including Ada, Healthily, Babylon, and Buoy Health [18]. As with a clinical evaluation standard, it will be essential that this process can keep pace with rapid development of digital products and does not become a barrier to innovation.

Protocols for Prospective Real-world Evidence

The clinical evaluation methods described in this paper relate to a theoretical validation of a model's performance that would often be performed by developers prior to product release or during the release of product updates. This should be distinguished from prospective clinical trials of OSCs in real-world settings. There is a strong need for studies of the real-world impact of OSCs on health care systems [7,14,23]. Robust prospective clinical studies comparing OSCs to existing provision, conducted by independent researchers, will be required in the 'preprimary care' and community setting to obtain a complete assessment of clinical product performance [27].

Some prospective clinical studies have been conducted to date comparing OSC triage to laypersons [28,29] and comparing OSC diagnoses to clinician diagnosis in real-world patients [30-32]. However, as with vignette-based evaluation studies, these prospective clinical studies demonstrate significant methodological variation, including the methods used for determining a gold standard outcome and benchmarking to this standard. Therefore, in addition to a standardized OSC

vignette-based clinical evaluation process, published protocols with standardized methods specific to prospective clinical studies would also be helpful. Conducting prospective trials at a pace that matches rapid iteration of products will present novel challenges and will require innovative approaches to evaluation methodology.

Guidance is also required on how to evaluate the extent to which OSCs' advice can be trusted and how user behavior varies compared to when they are given advice from health care professionals, such as triage nurses, pharmacists, or GPs [23]. Compliance with OSC advice is expected to be relatively low; evaluation of the NHS Pathways algorithm suggested that 30% of users who are told to attend emergency department using the algorithm do not comply. Conversely, 10.8% of users attend emergency department when they are advised against it [21].

User satisfaction as well as product usability and acceptability should also be further investigated. Some studies of usability of individual OSCs in real-world settings have been published [21,33], but further studies are required. This should include significant patient and public engagement and the exploration of differences among user sociodemographic groups that could impact health care inequality.

Conclusions

OSCs have significant potential to support the ability of individuals to self-care providing access to quality-assured health care information, and triaging recommendations. The use of these tools at scale could improve the rational use of scarce health resources, while also prompting patients with 'red flag' symptoms to seek emergency care promptly. However, there is currently no standardized way of clinically evaluating OSCs or benchmarking accuracy and safety. This makes comparison of OSC performance challenging and raises concerns about risks to patient safety and increasing health care system demand due to the use of OSCs. A set of objective guidelines for vignette-based clinical evaluation is required to instill confidence that an OSC is providing accurate and safe advice without adversely impacting health care systems.

The recommended requirements for a vignette-based OSC clinical evaluation standard summarized in Table 1 can help OSC developers, regulators, and health care systems work together to develop an effective validation standard. A clinical evaluation standard must be able to keep pace with the rapid iteration and development cycles of such technologies. Therefore, it will be essential that it is practical, pragmatic, and dynamic and does not introduce unnecessary barriers to innovation. The manual entry of vignettes that is often used in comparative studies is unlikely to be scalable, and therefore, the standard should also incorporate automated clinical evaluation methods.

The relative roles of vignette-based clinical evaluation versus prospective clinical studies will require further consideration. The rapid iteration of OSCs will likely make it unrealistic (due to both time and financial constraints) for prospective clinical studies to be conducted each time an OSC model is updated. Therefore, vignette-based evaluation is likely to continue to have a significant ongoing role in the validation of OSCs.

In future, the clinical evaluation of OSCs is expected to involve a mixture of vignette-based clinical evaluation by developers, third-party benchmarking, and prospective clinical studies. Therefore, alongside efforts to develop a clinical evaluation standard, the development of a third-party benchmarking process and the publication of protocols for prospective clinical studies to evaluate OSCs in real-world settings are of high priority.

Acknowledgments

BH and AEO are in part supported by the National Institute for Health and Care Research (NIHR) Applied Health Research (ARC) Northwest London. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care.

Authors' Contributions

AP conceptualized and drafted the manuscript. All authors reviewed, edited, and approved the manuscript.

Conflicts of Interest

BH is the Clinical Lead for Research and Development for eConsult, a platform for online consultations in primary, secondary, and urgent or emergency care. He has previously worked for Your.MD, an OSC provider (November 2019-May 2021). AP previously worked for Babylon Health, another OSC provider (June 2019-July 2020).

References

1. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun 11;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
2. Wallace W, Chan C, Chidambaram S. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. MedRxiv Preprint published online December 21, 2021. [doi: 10.1101/2021.12.21.21268167]
3. Iacobucci G. Row over Babylon's chatbot shows lack of regulation. BMJ 2020 Feb 28;368:m815. [doi: 10.1136/bmj.m815] [Medline: 32111647]

4. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021 Jul 15;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
5. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
6. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
7. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May 07;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
8. Guidance on Class I medical devices. UK Government. 2016. URL: https://www.gov.uk/government/collections/guidance-on-class-1-medical-devices [accessed 2022-08-08]
9. Evidence standards framework (ESF) for digital health technologies. NICE. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2022-08-10]
10. Digital triage in health services. Care Quality Comission. URL: https://www.cqc.org.uk/what-we-do/how-we-work-people/digital-triage-health-services [accessed 2022-08-08]
11. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
12. El-Osta A, Webber I, Alaa A, Bagkeris E, Mian S, Taghavi Azar Sharabiani M, et al. What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. BMJ Open 2022 Apr 27;12(4):e053566 [FREE Full text] [doi: 10.1136/bmjopen-2021-053566] [Medline: 35477872]
13. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
14. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. The Lancet 2018 Nov;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8]
15. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012 Feb 26;27(2):213-219 [FREE Full text] [doi: 10.1007/s11606-011-1804-8] [Medline: 21789717]
16. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994 Jun 23;330(25):1792-1796. [doi: 10.1056/nejm199406233302506]
17. Yager J, Linn LS, Leake B, Gastaldo G, Palkowski C. Initial clinical judgments by internists, family physicians, and psychiatrists in response to patient vignettes: I. Assessment of problems and diagnostic possibilities. Gen Hosp Psychiatry 1986 Jan;8(3):145-151. [doi: 10.1016/0163-8343(86)90072-1]
18. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 2000 Apr 05;283(13):1715-1722. [doi: 10.1001/jama.283.13.1715] [Medline: 10755498]
19. Noor P. Can we trust AI not to further embed racial bias and prejudice? BMJ 2020 Feb 12;368:m363. [doi: 10.1136/bmj.m363] [Medline: 32051165]
20. Turner J, O'Cathain A, Knowles E, Nicholl J, Tosh J, Sampson F, et al. Evaluation of NHS 111 pilot sites: final report. The University of Sheffield. 2012 Aug. URL: https://data.parliament.uk/DepositedPapers/Files/DEP2012-1694/HL3309-LibDoc.pdf [accessed 2022-10-13]
21. Lewis J, Stone T, Simpson R, Jacques R, O'Keeffe C, Croft S, et al. Patient compliance with NHS 111 advice: analysis of adult call and ED attendance data 2013-2017. PLoS One 2021 May 10;16(5):e0251362 [FREE Full text] [doi: 10.1371/journal.pone.0251362] [Medline: 33970946]
22. Turner J, O'Cathain A, Knowles E, Nicholl J. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. BMJ Open 2013 Nov 14;3(11):e003451 [FREE Full text] [doi: 10.1136/bmjopen-2013-003451] [Medline: 24231457]
23. Nguyen H, Meczner A, Burslam-Dawe K, Hayhoe B. Triage errors in primary and pre-primary care. J Med Internet Res 2022 Jun 24;24(6):e37209 [FREE Full text] [doi: 10.2196/37209] [Medline: 35749166]
24. Playing it safe - safety netting advice. MDU Journal. URL: https://mdujournal.themdu.com/issue-archive/issue-4/playing-it-safe---safety-netting-advice [accessed 2022-08-10]
25. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019 Oct 29;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]
26. Knitza J, Tascilar K, Gruber E, Kaletta H, Hagen M, Liphardt A, et al. Accuracy and usability of a diagnostic decision support system in the diagnosis of three representative rheumatic diseases: a randomized controlled trial among medical

students. Arthritis Res Ther 2021 Sep 06;23(1):233 [FREE Full text] [doi: 10.1186/s13075-021-02616-6] [Medline: 34488887]
27. Shah SS, Gvozdanovic A. Digital health; what do we mean by clinical validation? Expert Rev Med Devices 2021 Dec 12;18(sup1):5-8. [doi: 10.1080/17434440.2021.2012447] [Medline: 34845960]
28. Mills B, Hill M, Buck J, Walter E, Howard K, Raisinger A, et al. What constitutes an emergency ambulance call? Australas J Paramedicine 2019 Mar 22;16. [doi: 10.33151/ajp.16.626]
29. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475. [doi: 10.2196/24475]
30. Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. Arthritis Res Ther 2021 Apr 13;23(1):112 [FREE Full text] [doi: 10.1186/s13075-021-02498-8] [Medline: 33849654]
31. Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022 Jan 31;9(1):e32832 [FREE Full text] [doi: 10.2196/32832] [Medline: 35099395]
32. 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]
33. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]

Abbreviations

FG-AI4H: Focus Group on Artificial Intelligence for Health
GP: General Practitioner
NHS: National Health Service
NICE: National institute for Health and Care Excellence
OSC: online symptom checker

Edited by T Leung; submitted 19.02.22; peer-reviewed by J Knitza, M Schmieding, M Hill; comments to author 03.05.22; revised version received 15.09.22; accepted 11.10.22; published 26.10.22.

Please cite as:
Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A
Online Symptom Checkers: Recommendations for a Vignette-Based Clinical Evaluation Standard
J Med Internet Res 2022;24(10):e37408
URL: https://www.jmir.org/2022/10/e37408
doi:10.2196/37408
PMID:36287594

Viewpoint

Technical, Ethical, Legal, and Societal Challenges With Digital Twin Systems for the Management of Chronic Diseases in Children and Young People

David Drummond[1,2,3], MD, PhD; Adrien Coulet[2,4], PhD

[1]Department of Paediatric Pulmonology and Allergology, University Hospital Necker-Enfants Malades, AP-HP, Paris, France
[2]Heka Team, Centre de Recherche des Cordeliers, Inserm, Université Paris Cité, Paris, France
[3]Centre for Health Informatics, University of Manchester, Manchester, United Kingdom
[4]Inria Paris Research Centre, Paris, France

Corresponding Author:
David Drummond, MD, PhD
Department of Paediatric Pulmonology and Allergology
University Hospital Necker-Enfants Malades, AP-HP
149 rue de Sèvres
Paris, 75015
France
Phone: 33 144494848
Email: david.drummond@aphp.fr

Abstract

Advances in digital medicine now make it possible to use digital twin systems (DTS), which combine (1) extensive patient monitoring through the use of multiple sensors and (2) personalized adaptation of patient care through the use of software. After the artificial pancreas system already operational in children with type 1 diabetes, new DTS could be developed for real-time monitoring and management of children with chronic diseases. Just as providing care for children is a specific discipline—pediatrics—because of their particular characteristics and needs, providing digital care for children also presents particular challenges. This article reviews the technical challenges, mainly related to the problem of data collection in children; the ethical challenges, including the need to preserve the child's place in their care when using DTS; the legal challenges and the dual need to guarantee the safety of DTS for children and to ensure their access to DTS; and the societal challenges, including the needs to maintain human contact and trust between the child and the pediatrician and to limit DTS to specific uses to avoid contributing to a surveillance society and, at another level, to climate change.

(J Med Internet Res 2022;24(10):e39698) doi:10.2196/39698

KEYWORDS

artificial intelligence; pediatrics; medical cyber-physical systems; children; digital twin; child; personalized; cyber-physical; digital health; digital medicine; eHealth; ethics; legal; law; young people; youth; ethical; sensor; monitor; privacy; data collection; paediatric; pediatrician; paediatrician; chronic disease; medical system

Introduction

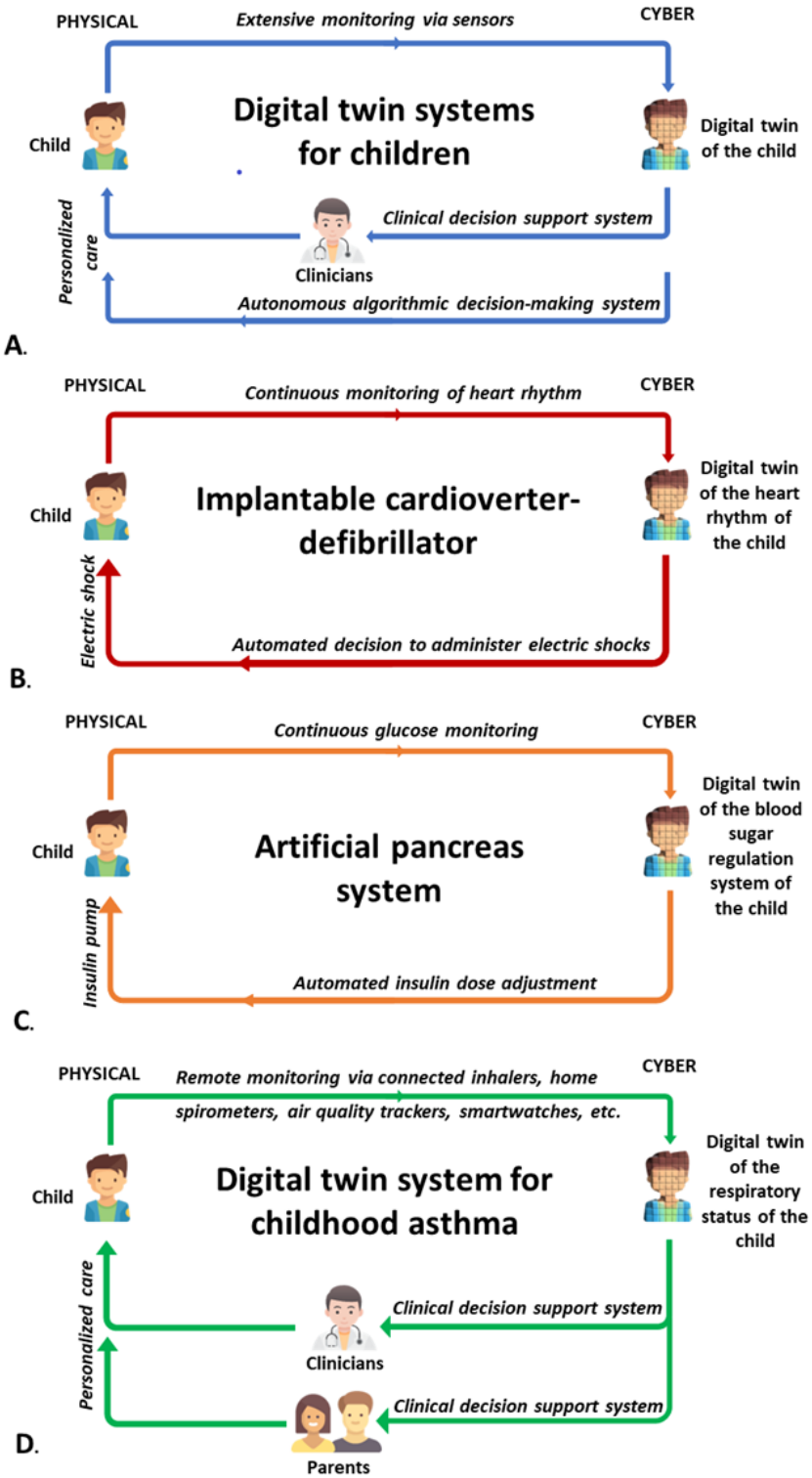
Throughout history, the practice of medicine has been constantly impacted by technological advances and societal developments. The first 2 industrial revolutions led to the development of new techniques for obtaining new information about the human body, resulting in the industrial collection of objective and quantitative data in the 20th century, including sensor-based physiological (eg, heart rate, oxygen saturation), biological, imaging, functional test, and increasingly “omics” (eg, genomics, proteomics) data [1]. From the 1970s onwards, the third industrial revolution, also known as the digital revolution, transformed this analogue data into digital data with 2 consequences in the first half of the 21st century: (1) the multiplication of computer models using artificial intelligence (AI) techniques to process patients’ health data and propose a diagnosis, establish a prognosis, and recommend a treatment [1,2] and (2) the possibility of obtaining, using the internet of things (IoT), a comprehensive representation of the patient's health status in real time (ie, a live digital replica of the patient, more commonly known as a “digital twin” [DT]) [3,4]. The combination of AI in DTs could lead to digital twin systems

(DTS), in which patients are constantly monitored from their homes and AI techniques adapt their care in real time.

DTS comprise a physical element—the patient—a cyber element—the patient's DT—and 2-way interactions between the physical and cyber elements: Sensors transform the patient's signal into the patient's DT, and software processes them to act through recommendations to the physician or automated adaptations on the patient's management [2] (Figure 1A). As the human body is extremely complex and its various mechanisms incompletely understood, it is currently not possible to perform a DTS of the whole human body. However, DTS of a function or an organ have gradually appeared. From the 1980s onwards, the first implantable cardioverter-defibrillators appeared: They collect the patient's heart rhythm in real time and, in the event of arrhythmia, automatically deliver an electric shock to restore a normal rhythm (Figure 1B). Such systems can be considered "pre-DTs" in that they originally measured only one parameter (heart rhythm) and the delivery of the electric shock was based on if-else algorithms. With a higher level of complexity, artificial pancreas systems, which combine blood glucose monitors, a virtual representation of the patient's physiology (interactions between measured blood glucose, physical activity, and diet), and actuators (delivery of the predicted optimal insulin dose via the insulin pump), have been developed for children with type 1 diabetes since the 2010s (Figure 1C). These artificial pancreas systems have been shown to improve clinical outcomes and quality of life for patients [5]. In the near future, it is likely that such systems will multiply for other chronic diseases in children, with an even higher degree of complexity. For example, in asthma, the most common chronic disease in children, it is necessary to take into account the multiple determinants of asthma symptoms, including treatment use (emergency and controller treatment) and the environment (eg, pollutants, allergens, weather conditions), to develop different computer models recommending the most appropriate controller treatment in real-time to health care professionals, the most appropriate mitigation measures in case of a high risk of asthma symptoms to families, and a way to involve children and families so that the recommendations made are followed at home. New connected objects (eg, connected inhalers, home spirometers, air quality trackers, smartwatches) and machine learning techniques now make it possible for such DTS to emerge in childhood asthma [6].

The challenges associated with the use of DTS have already been discussed [7-10]. However, just as providing care for CYP is a specific discipline—pediatrics—because of their particular characteristics and needs, providing digital care for CYP also presents particular challenges. Thus, the objective of this article was to review the technical, ethical, legal, and societal challenges associated with DTS for CYP.

Figure 1. Digital twin systems (DTS) for children and young people with chronic diseases: (A) sensors transform the child's "physical" information into "cyber" or digital data to adapt care, either by a health care professional or autonomously; (B) first example of DTS in childhood, which analyzes the heart rhythm in real time and autonomously delivers electric shocks; (C) more complex DTS, associating continuous monitoring of blood glucose, software adapting the insulin dose, and an insulin pump to deliver this dose; (D) may emerge in the near future for childhood diseases, such as asthma, requiring monitoring of many different determinants, machine learning techniques, and provision of recommendations to different actors (eg, children and their parents, teachers, and health care professionals).



Technical Challenges

The first requirement of digital medicine is digital data. However, both accessing and processing digital data are more complex for CYP than for adults.

DTS Without Smartphones

Smartphones have a central role in digital health by enabling (1) data collection through built-in functions (eg, GPS for geolocation) and interfaces for patient-reported outcome measures, (2) connection to the IoT via a Bluetooth connection,

and (3) patient feedback. In the best-equipped countries, smartphone ownership is limited to 25%, 50%, and 75% of children aged 8 years to 9 years, 10 years to 11 years, and 12 years to 13 years, respectively, complicating the use of digital health interventions for the remote management of children's diseases [11-14]. Furthermore, it is not intended to promote better smartphone equipment among CYP. In addition to the negative effects of screen time on children's cognitive and socioemotional development, smartphone addiction affects 1 in 4 CYP [15,16]. Thus, even if the DTS interfaces presented on smartphones do not pose an addiction problem, since they will be limited to supporting children in the daily management of their disease, providing a smartphone to children on this occasion could have negative effects on their development due to the other uses they would make of it (eg, games, social media). It is therefore preferable to use specific standalone devices to link the different connected objects as has been done for artificial pancreas systems. In the case of childhood asthma, a smartwatch that would both collect data relevant to monitoring the child's asthma control status (heart rate, oxygen saturation, activity) and provide appropriate and timely recommendations on its screen without allowing other types of internet access would be an interesting solution.

Designing Devices for Growing CYP

All parents have experienced the recurrent changes in size of clothing and shoes as children grow. A child's height doubles in the first 4 years, from an average of 50 cm to 100 cm, and then increases again by at least 1.5 times over the next 15 years. Similarly, a child's abilities develop impressively, from the infant who does not yet walk and talk to the adolescent capable of the most extreme sports and complex reasoning. Finally, in the medical field, physiological values vary constantly with the age of the child (eg, a heart rate of 50 bpm is normal for a teenager but abnormally low for an infant), as do the expected results of additional tests. This complicates the task for manufacturers of connected devices who need to provide different model sizes, develop appropriate interfaces for children and adolescents of different ages, and adapt the standards for physiological parameters according to the age of the child. These adaptations are not without risk: In the field of cardiac medical devices, an attempt to miniaturize an implantable cardioverter defibrillator resulted in a higher risk of failure in young patients [17]. In the case of an asthma DTS providing recommendations for children, we shall ensure that the recommendations given are age-appropriate, using oral instructions for a 6- or 7-year-old rather than text messages, which would be appropriate for teenagers.

Protecting Children From Devices

The use of connected objects to collect data from children poses particular risks. Similar to a toy with a defective design, children may choke on, or simply ingest, small parts that may come off the object; they can also be exposed to chemicals that are carcinogenic or mutagenic or to endocrine-disrupting substances. Thus, devices intended to collect data from children's bodies must be manufactured taking into account the additional risks they may pose to children, which is another challenge compared with manufacturing devices for adults.

Protecting Devices From Children

In the other direction, children pose specific risks to the connected objects. In the same way that children regularly come home having lost or broken their glasses, connected objects are more likely to be lost or broken when used with children than with adults. This is the consequence of their age-related activities but also of a less cautious attitude toward their personal belongings. For example, connected inhalers that automatically record children's use of asthma medication were reported lost or damaged by up to 50% of families [18,19].

Developing Models for CYP

CYP pose particular problems when it comes to creating models with traditional supervised approaches from real-world data. The amount of data available is lower, due to the young age of patients (less historical data), logistical and legal difficulties in obtaining the data, and the lower prevalence of diseases in children than in adults in high-income countries [20]. Furthermore, the heterogeneity of the pediatric population is such that it is unsound to learn a single model and necessary to split data in several subgroups. This leads to smaller data sets, which makes it more difficult to obtain performant models for CYP. Fine-tuning models developed for adults may be an option to overcome these limitations, but to our knowledge, it has not yet been applied to create pediatric models.

Ethical Challenges

The main ethical question related to the use of digital health in pediatrics is whether the use of DTS is in the best interests of the child. The notion of "best interests of the child" is derived from Article 3 of the United Nations Convention on the Rights of the Child [21]. This is a deliberately ill-defined concept, which needs to be assessed on a case-by-case basis, but several principles that can be used to guide decision-makers have been provided [22]. In this section, we consider how the adoption of digital health for CYP may pose a threat to these "best interests" in light of the 4 principles of biomedical ethics identified by Beauchamp and Childress [23]: respect for autonomy, beneficence, non-maleficence, and justice (Figure 2).

Figure 2. Potential benefits and risks of pediatric digital twin systems (DTS) at the individual level, which require a specific premarket assessment that takes into account not only the health impacts but also the impact of DTS on the child's physical, mental, or social development. AI: artificial intelligence; PROM: patient-reported outcome measure.

Potential benefits

- **Improved health**
  - Increased autonomy in life (new activities allowed)
  - Improved mental well-being (decreased cognitive load)
  - Improved social development (engagement in new social activities)
  - Increased participation in sports (decreased symptoms during exercise)
- **Improved access to health care**
  - Remote disease management
  - Timely treatment adaptation
  - Standardization of the level of care

Potential threats

- **Decreased health**
  - Risk of toxic exposure, choking (hardware)
  - Risk of malfunction (software)
- **Decreased autonomy in care**
  - Decreased participation in symptom description (child's PROM ignored)
  - Decreased participation in daily management (complexity)
- **Decreased well-being**
  - Decreased engagement in recreation and sports (complexity)
  - Reduced participation in social activities (shame and/or complexity of devices)
- **Exclusion of health care**
  - Not affordable
  - Not usable by families with low levels of digital literacy
  - Self-exclusion of low-income families (no trust in AI)

Preserving the Development of Autonomy

Autonomy means being governed by oneself, in thought and action [24]. As the principle of autonomy cannot be applied as such to children, who are inherently dependent on adults to meet their needs, it was proposed instead to protect the development of the child's autonomy [25]. Pediatricians encourage the development of autonomy in children with diseases by encouraging them to describe their symptoms themselves and by supporting them in taking responsibility for their own care.

Preserving Children's Participation in Describing Their Symptoms

Pediatricians encourage children to talk about their symptoms themselves to recognize their place as individuals in the consultation but also because their statements are frequently more accurate than those of their parents [26]. A first risk of DTS is that they do not take into account patient-reported outcome measures from children. Indeed, although it is possible for pediatricians to detect when a child is hesitant in their response or gives a fanciful answer, the direct entry of data by the child into a digital collection system does not allow this assessment. In addition, if the child enters incorrect information, the DTS is likely to provide an incorrect and potentially dangerous action or answer for the child's health. In this context, developers may choose not to include the answers entered by children or ask parents to validate all entries. In both cases, this would be a step backwards in pediatric practice to a time when the child's voice was not taken into account. A second risk of DTS may arise if there is a discrepancy between the child's reported symptoms and the DTS assessment. In this case, adults may tend to believe the DTS rather than the child. Indeed, between 2 contradictory pieces of information, one given by a human and the other by a computer system, humans tend to believe the information given by the computer system [27]. For example, in medicine, clinicians override their own correct decisions in favor of incorrect advice from a decision support system in 6%-11% of cases.[28]. Since children may be perceived as having little credibility [29], this automation bias is likely to be exacerbated in this situation, with the DTS becoming the reference point for adults on the child's condition. For example, consider the case of an 8-year-old asthmatic boy who calls his parents at work to tell them that he is starting to have trouble breathing. The parents check the child's DT status on a dedicated mobile app, note that the risk of an asthma attack is very low, and explain to their child that no, he is not having an asthma attack. How will this child react to the fact that the symptoms he is reporting are not being heard, regardless of the reality or not of the asthma attack? One possibility is that he will lose confidence in himself and his feelings, as a computer system would be more trusted by his parents than his own word.

Preserving Children's Empowerment in Their Daily Care

Currently, CYP gradually become responsible for the day-to-day management of their disease, in agreement with their parents and pediatricians. This process can be delayed if DTS are complex to use, whether the complexity is due to the hardware or software components of the DTS. Conversely, DTS can also increase children's autonomy in managing their daily care if they rely on easy-to-use devices and software: In the case of type 1 diabetes, most CYP and parents reported greater control and autonomy in managing their diabetes with insulin pumps after an initial learning period than with injections [30].

Beneficence

Beneficence refers to the responsibility of professionals to promote the well-being of their patients [31]. It is clear that, before being used in children, DTS must be evaluated for both efficacy and safety through appropriate clinical studies. However, DTS may also pose particular threats to young people and their development.

Preserving the Need for Children to Engage in Play and Recreational Activities

By improving their health, DTS are expected to enable children to participate more in play and recreational activities. However, DTS can also prevent children from playing or participating in certain sports: Children using the artificial pancreas system for their type 1 diabetes have reported difficulties when playing sports due to the tendency of the insulin pump to fall [30]. Using a smartwatch for children with asthma may also prevent them from swimming if it has not been anticipated that the device should be waterproof. The right of the child to engage in play and recreational activities appropriate to his or her age is enshrined in the UN Convention on the Rights of the Child (Art. 31), and DTS shall not require children to behave like adults.

Preserving Children's Mental Well-being

Children need to be protected from chronic stress in order to be in the best condition to develop their ability to control their emotions, focus on tasks, and form healthy social relationships [32,33]. DTS can help to reduce this chronic stress through better control of the disease and delegation of decision-making. However, there is a risk that they place the child under continuous pressure. At present, a child with asthma may feel stressed when approaching a doctor's appointment because he knows that he has not taken his medication properly and that his pediatrician is going to re-explain the importance of it. However, in between visits, he is not being monitored and therefore not under the stress of being held accountable. For example, children with asthma may feel under constant pressure with the continuous monitoring introduced by connected inhalers to track their adherence and the associated alerts to their physician if they forget to take their medication [19]. The preservation of the child's physical health must be balanced with their mental well-being.

Preserving Children's Social Development

Children need to interact with their peers, including forming strong friendships, to develop their social skills [34]. Again, DTS can protect these peer interactions by improving the health of CYP and promoting school attendance and participation in various activities. However, DTS can also impair the social life of CYP in 2 ways. First, the complexity of DTS may prevent parents from leaving their child with other adults. In the case of type 1 diabetes, the complexity of the artificial pancreas system and difficulty of explaining how it works were perceived by parents as barriers to leaving children with other people, for example to spend a night at a friend's house [30]. Second, CYP who have to wear their DTS device on their body feel different from others, which may prevent them from reaching out to others. Adolescent girls with type 1 diabetes expressed how the insulin pump could make them feel different about their bodies and how they tried to hide it for fear of not being accepted by their peers [35]. Similarly, children with a smartwatch or wearable air quality tracker to continuously monitor their asthma could be singled out by their peers.

Justice

The principle of justice refers to the delivery of equal treatment and care according to the particular patient's needs, as well as just allocation of available resources [36]. Providing appropriate care for each child according to their needs is all the more important as poor child health limits the potential and development of children, leading to reduced health and life chances in adulthood [37,38]. Yet, the main risk of DTS is that it exacerbates inequalities in health care for children. First, disadvantaged families may not be able to afford DTS, thereby excluding their children from the most effective (if proven) treatment strategies. Second, families with low levels of digital literacy may find it more difficult to use DTS, resulting in less effective treatment adaptations. The third risk is that of self-exclusion of disadvantaged families. Several studies found that lower levels of education, lower levels of employment, and lower household income are associated with negative views and reluctance to participate in research programs involving AI [39-42], raising the risk that algorithms will be trained on data from advantaged families and optimized for these populations. However, DTS can also contribute to reducing inequalities in children's care; for example, by standardizing the care of children and bypassing the doctor for certain decisions, children from disadvantaged families, living in remote areas or with out-of-date doctors, can receive the same care with DTS as children living in privileged areas. For example, the use of connected inhalers that automatically send alerts to health care professionals in case of an asthma attack has been shown to particularly improve the care of the most disadvantaged children [19].

In conclusion, ethical dilemmas can arise in many ways when developing a DTS for CYP, given the need to preserve the health of the child; preserve their physical, mental, and social development; and be anticipated.

Legal Challenges

As stated by the policy guidance on AI for children from the United Nations International Children's Emergency Fund, children need protection from AI (do no harm) and provision of effective AI systems (do good) [43]. In health, this translates to the need to protect them from the risks posed by DTS and to ensure that children are offered DTS to improve their health.

Protecting Children From Specific Risks: Do No Harm

Children need special protection from DTS risks for 2 reasons: They have special characteristics that expose them to greater or potentially different risks in comparison with adults, and they are developing beings who may face dramatic consequences for the rest of their lives if their health is adversely affected. Because of these specificities, the approval of a drug for children is subject to a specific process in the United States and European Union. Sponsors are required to provide a pediatric study plan to the Food and Drug Administration (FDA) in the United States or its equivalent pediatric investigation plan to the European Medicines Agency (EMA), which are reviewed by specific committees taking in account pediatric considerations: the FDA's Pediatric Review Committee in the United States and the EMA's Paediatric Committee in the European Union [44].

DTS are not subject to this regulation. Under the current regimes, they are considered medical devices and not medicines.

The regulation of medical devices, although recently strengthened in the European Union by the regulation 2017/745 [45], is not as strict as that of medicines from a pediatric perspective [46]. In the United States, the FDA has issued specific guidance for the development of pediatric medical devices [47], but there is no dedicated pediatric committee to assess how these devices are evaluated or adapted for children as there is for medicines and as advocated by the American Academy of Pediatrics [48]. In the European Union, the regulation is limited to stating that the presence of carcinogenic, mutagenic, reprotoxic, or endocrine-disrupting substances must be justified for devices intended for children [45]. Because medical devices are no less risky than medicines, it is important that a similar pathway with pediatric investigation and study plans, reviewed by specialized pediatric committees, is established by the legislators for the approval of pediatric DTS.

Ensuring CYPs Are Provided With Appropriate DTS: Do Good

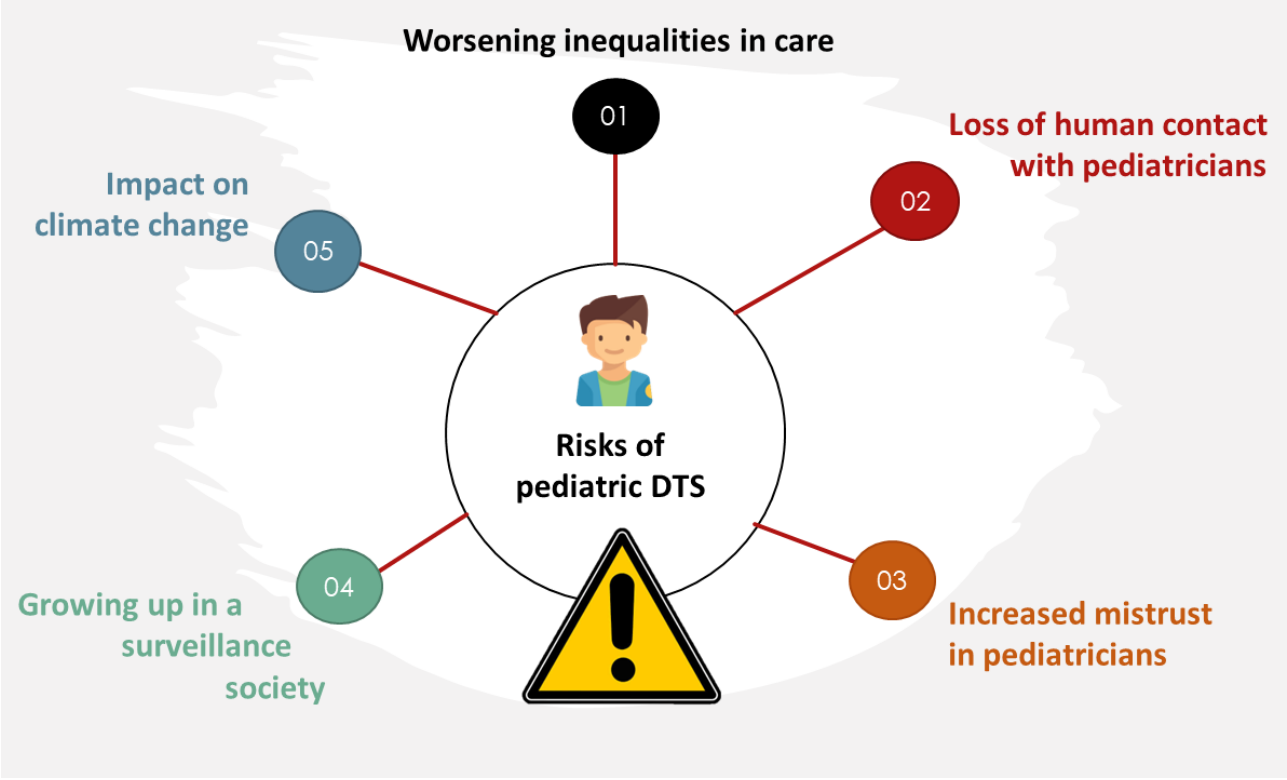
If legislators must protect children from the risks that a DTS could pose, they must also protect children from the risk of not having access to DTS that could improve their health. Industry has consistently been less interested in developing drugs for children because of the small number of children affected, increased regulatory constraints, and difficulty of conducting clinical trials in this population, and this trend continues with medical devices including DTS [49-52]. As an example, if multiple algorithms have been used by manufacturers of implantable cardioverter-defibrillators to correctly diagnose atrial and ventricular arrhythmias, these have not been tested in children [17]. For drugs, the situation improved with the introduction of regulations in the European Union and the United States with 2 objectives: (1) to compel pharmaceutical companies to include studies with pediatric populations if the disease or condition for which the drug is indicated occurs in children and (2) to encourage these same companies to carry out these studies by providing a financial incentive in the form of an extension of intellectual property rights [49]. However, such schemes do not exist for medical devices. In the United States, the FDA has launched a series of initiatives to encourage manufacturers to develop pediatric medical devices, with some results [53]. Such initiatives are awaited in Europe.

In conclusion, the risk is that, in the absence of legislative constraints or incentives, market forces will continue to widen the gap between the quantity and quality of DTS developed for adults and children.

Societal Challenges

DTS in pediatrics is currently limited to a few indications (automated adaptation of insulin doses by artificial pancreas in type 1 diabetes, automated detection and treatment of arrythmia by implantable cardioverter-defibrillators), and pediatricians prescribing DTS are unlikely to feel they are participating in societal change. However, in combination with other digital transformations in society, DTS are likely to lead to profound changes in our lifestyles with important consequences on children. Apart from the risk of exacerbating social inequalities already detailed, DTS can contribute to several societal challenges (Figure 3).

Figure 3. Risks of pediatric digital twin systems (DTS) at the societal level.



Preserving Human Contact in Care

If most medical decisions are made autonomously and remotely by DTS, this would result in fewer in-person visits to pediatricians, in line with a more general trend of reduced face-to-face interactions in society [54]. The risk of dehumanization associated with DTS may be a greater problem for pediatricians, who made the choice of a particularly "people-oriented" specialty, in contrast to "technology-oriented" specialties (eg, anesthesiology, radiology) [55]. The risk is that an increasingly technological approach to child health care, combined with a reduction in face-to-face interactions, will lead many to a loss of meaning in practice and to burnout. This may result in a gradual replacement of these doctors by professionals proficient with information technology but less skilled in human relations.

Preserving Trust

The use of DTS in pediatrics may contribute to a growing distrust of pediatricians. The digitalization of the world, whether through access to medical information on the Internet or conspiracy groups on social networks, has already increased parents' distrust of their pediatricians [56,57]. Families may have less confidence in their pediatrician if they feel that the latter is clearly not mastering the DTS or if the recommendations made by the DTS and the pediatrician differ. In the latter case, explainable models would help the pediatrician to justify disagreement with the DTS and maintain trust. An additional difficulty is that most DTS will aim to prevent a complication or exacerbation from occurring and therefore act while the patient has few or no symptoms. The positive effect of DTS will therefore be less perceptible (absence of an uncertain event) than an exceptional side effect highlighted in social networks. For example, a DTS for childhood asthma may prevent thousands of asthma attacks by providing appropriate recommendations to families when the risk of an asthma attack is predicted for the coming week but be blamed for an undetected severe asthma attack. Finally, pediatricians are perceived to be acting in the best interests of the child, whereas this may not be the case for the companies behind DTS, increasing mistrust [58].

Preventing Children From Growing Up in a Surveillance Society

The use of DTS requires the collection of a large amount of data, which are increasingly being collected continuously from the child's home or even body. This constant monitoring of children to ensure their safety is a general trend in our societies, in the same way that the proliferation of surveillance cameras was intended to ensure the safety of the population. It can lead to a surveillance society where everyone watches each other, with children being a main target because of their vulnerability. Parents could demand to monitor their children at all times for their safety, since the data are already available, being continuously collected for health purposes. Pediatricians may also ask to use indirect information about parents' behavior collected from the home devices of the child to monitor the risk of child abuse. If such a surveillance society was to emerge, children would grow up in an optimized state of physical health but would probably be more anxious, dependent, and conventional in adulthood, unable to make decisions on their own values [59,60]. Again, legislation implemented by each society will be crucial in ensuring that the best interest of the child is taken into account, at both the individual and societal levels.

Preserving the Environment

Climate change is one of the major challenges of our century. Young people understand this and are actively campaigning for policies to reduce carbon emissions [61]. CYP are indeed the first to be affected by the consequences of global warming: not only will children born today live in a world with a temperature 4 degrees higher than the pre-industrial era but the effects of global warming on health are much more significant for children than for adults [62]. For example, with regard to the development of diseases favored by global warming, 93% fall upon children [63].

Among the sectors responsible for greenhouse gas emissions, the digital sector is growing in importance every year, increasing in the contribution to global emissions from 2.5% to 3.7% between 2013 and 2019 [64]. Medical devices are the first part of the problem. They usually require raw materials such as rare metals, and most of them are single-use devices with limited recyclability [65]. The flow and storage of the data generated by these devices in data centers are another part of the problem. Currently, data centers account for 1% of total global electricity demand, with about one-half of this energy being used to cool servers [66,67]. Finally, it was recently shown that training a single AI model could generate CO_2 emissions equivalent to those of a passenger making 300 flights between New York and San Francisco [68]. Thus, even with efforts to reduce electricity consumption in data centers and to move toward "green AI" with efficient models, the general use of DTS would generate a significant amount of greenhouse gas emissions [69]. Conversely, if DTS are effective and improve children's health, they would reduce the use of health care (eg, travel to hospitals, hospital admissions) and in turn would reduce the associated amount of greenhouse gas emissions. Thus, it is currently essential to carry out studies of the environmental impact of digital health interventions, in addition to efficacy trials and medicoeconomic studies, as envisaged by the National Institute for Health and Care Excellence in the United Kingdom [70]. This will participate in safeguarding the future of children.

Conclusion

The use of DTS for children poses specific challenges at individual and societal levels (Figure 4).

Since the values at stake are at different levels (preserving the child's life, preserving the child's quality of life and development, preserving life in society, preserving the planet), the ethical approach that seems most appropriate when developing and evaluating a DTS is that of "value pluralism" [71]. This approach recognizes many different, equally fundamental moral values, which may conflict with each other without a predominant value. Indeed, improving children's health is as important as ensuring their quality of life and future development, promoting a society in which they can flourish

and leaving them with a livable planet. To take into account all these dimensions, the development of DTS needs to involve many stakeholders at all stages, from the development phase to the evaluation phase (Figure 5).

The evaluation of a pediatric DTS must balance the expected effects on the child's health and its beneficial consequences (increased autonomy, well-being, socialization) against the risks posed by the DTS, whether individual (risk of exposure to toxic substances, stigmatization), societal (contribution to increased inequalities, surveillance society), or global (climate change). This specific evaluation should be supported by specific legislation on pediatric DTS and by incentives by governments and private foundations to promote children's access to DTS. Indeed, children should not be deprived of DTS, which, if effective, could be a real game changer in the management of their diseases.

Figure 4. Impact of digital twin systems in pediatrics at different levels and links between the different values. Green arrows indicate a positive impact, and red arrows indicate a negative impact.

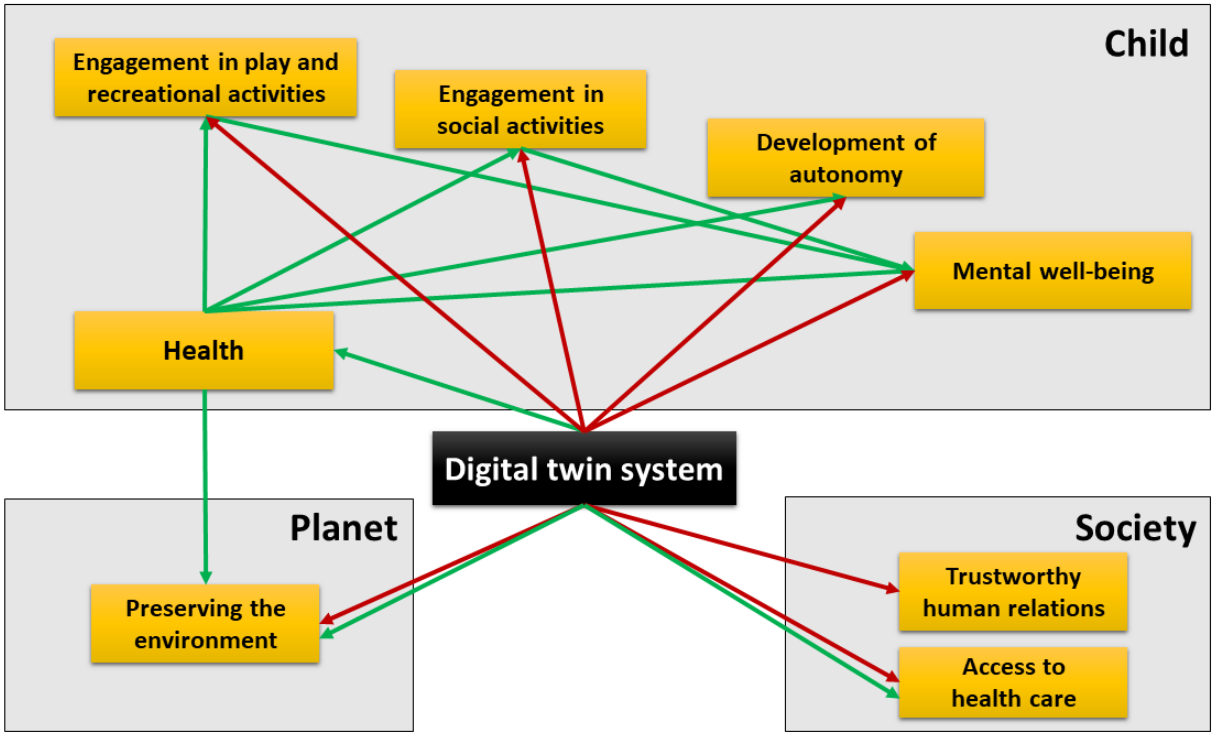
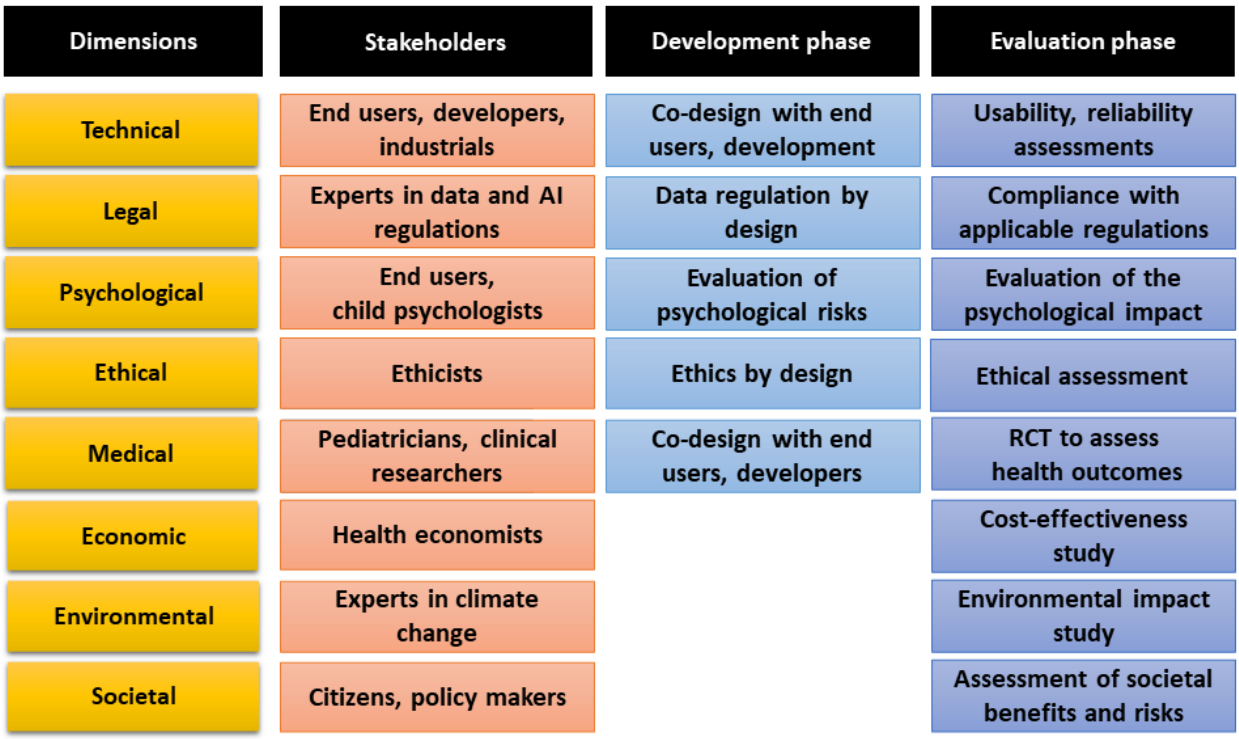


Figure 5. Dimensions of digital twin systems for pediatrics, stakeholders involved, and stakeholders' roles during the development and evaluation phases. AI: artificial intelligence; RCT: randomized controlled trial.



Acknowledgments

David Drummond is supported by a grant from the Société Française de Pneumologie et Allergologie Pédiatriques – AstraZeneca.

Conflicts of Interest

None declared.

References

1. Drummond D. Between competence and warmth: the remaining place of the physician in the era of artificial intelligence. NPJ Digit Med 2021 May 14;4(1):85 [FREE Full text] [doi: 10.1038/s41746-021-00457-w] [Medline: 33990682]
2. Exarchos K, Beltsiou M, Votti C, Kostikas K. Artificial intelligence techniques in asthma: a systematic review and critical appraisal of the existing literature. Eur Respir J 2020 Sep;56(3):a [FREE Full text] [doi: 10.1183/13993003.00521-2020] [Medline: 32381498]
3. Popa EO, van Hilten M, Oosterkamp E, Bogaardt M. The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks. Life Sci Soc Policy 2021 Jul 05;17(1):6 [FREE Full text] [doi: 10.1186/s40504-021-00113-x] [Medline: 34218818]
4. Laubenbacher R, Sluka JP, Glazier JA. Using digital twins in viral infection. Science 2021 Mar 12;371(6534):1105-1106 [FREE Full text] [doi: 10.1126/science.abf3370] [Medline: 33707255]
5. Liu Z, Meyendorf N, Mrad N. The role of data fusion in predictive maintenance using digital twin. AIP Conference Proceedings 2018;1949(1):020023. [doi: 10.1063/1.5031520]
6. Drummond D. [Digital tools for remote monitoring of asthma patients: Gadgets or revolution?]. Rev Mal Respir 2022 Mar;39(3):241-257. [doi: 10.1016/j.rmr.2022.01.018] [Medline: 35219562]
7. Gerke S, Minssen T, Cohen G. Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare. Cambridge, MA: Academic Press; 2020:295-336.
8. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA 2019 Dec 24;322(24):2377-2378. [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
9. Lee I, Sokolsky O, Chen S, Hatcliff J, Jee EY, Kim BG, et al. Challenges and research directions in medical cyber–physical systems. Proc. IEEE 2012 Jan;100(1):75-90. [doi: 10.1109/JPROC.2011.2165270]
10. Recchia G, Maria Capuano D, Mistri N, Verna R. Digital therapeutics-what they are, what they will be. Act Scie Medic 2020 Feb 26;4(3):01-09. [doi: 10.31080/asms.2020.04.0575]

11. Koptyug E. Smartphone ownership among children in Germany in 2020, by age group. Statista. 2022 Jan 13. URL: https://www.statista.com/statistics/1283678/smartphone-children-owners-age-group-germany/ [accessed 2022-10-24]
12. Smartphone penetration rate by age group in France 2020. Statista. 2022 Feb 08. URL: https://fr.statista.com/statistiques/505110/taux-de-penetration-du-smartphone-par-age-france/ [accessed 2022-10-24]
13. Smartphone ownership in the U.S. 2015-2021, by age group. Statista. 2022 Oct 18. URL: https://www.statista.com/statistics/489255/percentage-of-us-smartphone-owners-by-age-group/ [accessed 2022-10-24]
14. The Common Sense Census: Media Use by Tweens and Teens, 2019. Common Sense Media. 2019 Oct 28. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2019 [accessed 2022-10-24]
15. Sohn S, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019 Nov 29;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
16. Domingues-Montanari S. Clinical and psychological effects of excessive screen time on children. J Paediatr Child Health 2017 Apr 06;53(4):333-338. [doi: 10.1111/jpc.13462] [Medline: 28168778]
17. Dubin AM, Cannon BC, Saarel EV, Triedman JK, Berul CI, Bar-Cohen Y, et al. Pediatric and congenital electrophysiology society initiative on device needs in pediatric electrophysiology. Heart Rhythm 2019 Apr;16(4):e39-e46. [doi: 10.1016/j.hrthm.2018.12.021] [Medline: 30579977]
18. Morton RW, Elphick HE, Rigby AS, Daw WJ, King DA, Smith LJ, et al. STAAR: a randomised controlled trial of electronic adherence monitoring with reminder alarms and feedback to improve clinical outcomes for children with asthma. Thorax 2017 Apr;72(4):347-354. [doi: 10.1136/thoraxjnl-2015-208171] [Medline: 27815524]
19. Gupta RS, Fierstein JL, Boon KL, Kanaley MK, Bozen A, Kan K, et al. Sensor-based electronic monitoring for asthma: a randomized controlled trial. Pediatrics 2021 Jan;147(1):1 [FREE Full text] [doi: 10.1542/peds.2020-1330] [Medline: 33386336]
20. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
21. Convention on the Rights of the Child. United Nations Treaty Collection. 1989 Nov 20. URL: https://treaties.un.org/pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-11&chapter=4&clang=_en [accessed 2022-10-24]
22. General comment No. 14 (2013) on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1). United Nations. 2013. URL: https://www2.ohchr.org/english/bodies/crc/docs/gc/crc_c_gc_14_eng.pdf [accessed 2022-10-24]
23. Beauchamp TL, Childress JF. Principles of biomedical ethics. Oxford, England: Oxford University Press; 2001.
24. Kamii C. Autonomy: the aim of education envisioned by Piaget. The Phi Delta Kappan 1984;65(6):410-415 [FREE Full text]
25. Baines P. Medical ethics for children: applying the four principles to paediatrics. J Med Ethics 2008 Mar;34(3):141-145. [doi: 10.1136/jme.2006.018747] [Medline: 18316452]
26. Davis K, Disantostefano R, Peden D. Is Johnny wheezing? Parent-child agreement in the Childhood Asthma in America survey. Pediatr Allergy Immunol 2011 Feb;22(1 Pt 1):31-35. [doi: 10.1111/j.1399-3038.2010.01016.x] [Medline: 21261742]
27. Cummings M. Automation Bias in Intelligent Time Critical Decision Support Systems. 2004 Presented at: AIAA 1st Intelligent Systems Technical Conference; September 20-22, 2004; Chicago, IL URL: https://arc.aiaa.org/doi/10.2514/6.2004-6313 [doi: 10.2514/6.2004-6313]
28. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127 [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
29. Baumtrog M, Peach H. They can't be believed: children, intersectionality, and epistemic injustice. Journal of Global Ethics 2019 Dec 20;15(3):213-232. [doi: 10.1080/17449626.2019.1695280]
30. Alsaleh FM, Smith FJ, Thompson R, Al-Saleh MA, Taylor KMG. Insulin pump therapy: impact on the lives of children/young people with diabetes mellitus and their parents. Int J Clin Pharm 2014 Oct;36(5):1023-1030. [doi: 10.1007/s11096-014-9990-1] [Medline: 25108411]
31. Singh JP, Ivory M. Beneficence/Nonmaleficence. In: The Encyclopedia of Clinical Psychology. Hoboken, NJ: John Wiley & Sons Inc; Jan 23, 2015.
32. Thompson RA. Stress and child development. Future Child 2014;24(1):41-59. [doi: 10.1353/foc.2014.0004] [Medline: 25518702]
33. Shonkoff JP, Garner AS, Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, Section on Developmental and Behavioral Pediatrics. The lifelong effects of early childhood adversity and toxic stress. Pediatrics 2012 Jan;129(1):e232-e246. [doi: 10.1542/peds.2011-2663] [Medline: 22201156]
34. Ladd GW, Burgess KB. Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? Child Dev 2001 Sep;72(5):1579-1601. [doi: 10.1111/1467-8624.00366] [Medline: 11699688]

35. Ritholz M, Smaldone A, Lee J, Castillo A, Wolpert H, Weinger K. Perceptions of psychosocial factors and the insulin pump. Diabetes Care 2007 Mar;30(3):549-554. [doi: 10.2337/dc06-1755] [Medline: 17327319]
36. Halvorsen K, Førde R, Nortvedt P. The principle of justice in patient priorities in the intensive care unit: the role of significant others. J Med Ethics 2009 Aug;35(8):483-487. [doi: 10.1136/jme.2008.028183] [Medline: 19644006]
37. Wickham S, Anwar E, Barr B, Law C, Taylor-Robinson D. Poverty and child health in the UK: using evidence for action. Arch Dis Child 2016 Aug;101(8):759-766 [FREE Full text] [doi: 10.1136/archdischild-2014-306746] [Medline: 26857824]
38. Basch C. Healthier students are better learners: a missing link in school reforms to close the achievement gap. J Sch Health 2011 Oct;81(10):593-598. [doi: 10.1111/j.1746-1561.2011.00632.x] [Medline: 21923870]
39. Zhang B, Dafoe A. Artificial intelligence: American attitudes and trends. SSRN Journal 2019:1. [doi: 10.2139/ssrn.3312874]
40. Cameron D, Maguire K. Public views of machine learning: Digtial natives. The Royal Society. 2017 Oct. URL: https://royalsociety.org/-/media/policy/projects/machine-learning/digital-natives-16-10-2017.pdf [accessed 2022-10-24]
41. Attitudes towards the impact of digitisation and automation on daily life. European Commission. 2017 May 10. URL: https://digital-strategy.ec.europa.eu/en/news/attitudes-towards-impact-digitisation-and-automation-daily-life [accessed 2022-10-24]
42. Abdoul C, Cros P, Coutier L, Hadchouel A, Neuraz A, Burgun A, et al. Parents' views on artificial intelligence for the daily management of childhood asthma: a survey. J Allergy Clin Immunol Pract 2021 Apr;9(4):1728-1730.e3. [doi: 10.1016/j.jaip.2020.11.048] [Medline: 33290917]
43. Dignum V, Penagos M, Pigmans K, Vosloo S. Policy guidance on AI for children. UNICEF. 2021 Nov. URL: https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children [accessed 2022-10-24]
44. Penkov D, Tomasi P, Eichler I, Murphy D, Yao LP, Temeck J. Pediatric medicine development: an overview and comparison of regulatory processes in the European Union and United States. Ther Innov Regul Sci 2017 May 01;51(3):360-371 [FREE Full text] [doi: 10.1177/2168479017696265] [Medline: 28674673]
45. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.). European Union. 2017 May 05. URL: http://data.europa.eu/eli/reg/2017/745/oj/eng [accessed 2022-10-24]
46. Huijghebaert S, De Bruyne P, Allegaert K, Vande Velde S, De Bruyne R, Van Biervliet S, et al. Medical devices that look like medicines: safety and regulatory concerns for children in Europe. Arch Dis Child 2020 Feb;105(2):147-154. [doi: 10.1136/archdischild-2018-316391] [Medline: 31533915]
47. Premarket Assessment of Pediatric Medical Devices. Food & Drug Administration. 2014 Mar. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/premarket-assessment-pediatric-medical-devices [accessed 2022-10-24]
48. Pediatric Medical Product Innovation. American Academy of Pediatrics. 2020 Nov 10. URL: http://www.aap.org/en/advocacy/transition-plan-2020/leading-nation/pediatric-medical-product-innovation/ [accessed 2022-10-24]
49. Ward RM, Benjamin DK, Davis JM, Gorman RL, Kauffman R, Kearns GL, et al. The need for pediatric drug development. J Pediatr 2018 Jan;192:13-21 [FREE Full text] [doi: 10.1016/j.jpeds.2017.08.011] [Medline: 28942898]
50. Hwang TJ, Kesselheim AS, Bourgeois FT. Postmarketing trials and pediatric device approvals. Pediatrics 2014 May;133(5):e1197-e1202 [FREE Full text] [doi: 10.1542/peds.2013-3348] [Medline: 24733871]
51. Espinoza JC. The scarcity of approved pediatric high-risk medical devices. JAMA Netw Open 2021 Jun 01;4(6):e2112760 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12760] [Medline: 34156458]
52. Lee SJ, Cho L, Klang E, Wall J, Rensi S, Glicksberg BS. Quantification of US Food and Drug Administration premarket approval statements for high-risk medical devices with pediatric age indications. JAMA Netw Open 2021 Jun 01;4(6):e2112562 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12562] [Medline: 34156454]
53. FDA's Efforts to Optimize Medical Device Innovation for Pediatrics. Food & Drug Administration. URL: https://www.fda.gov/media/115027/download [accessed 2022-10-24]
54. Twenge JM, Spitzberg BH. Declines in non‐digital social interaction among Americans, 2003–2017. J Appl Soc Psychol 2020 Apr 10;50(6):363-367. [doi: 10.1111/jasp.12665]
55. Hojat M, Gonnella JS, Mangione S, Nasca TJ, Magee M. Physician empathy in medical education and practice: experience with the Jefferson scale of physician empathy. Seminars in Integrative Medicine 2003 Mar;1(1):25-41. [doi: 10.1016/s1543-1150(03)00002-4]
56. Sood N, Jimenez DE, Pham TB, Cordrey K, Awadalla N, Milanaik R. Paging Dr. Google: the effect of online health information on trust in pediatricians' diagnoses. Clin Pediatr (Phila) 2019 Jul;58(8):889-896. [doi: 10.1177/0009922819845163] [Medline: 31043059]
57. Stahl J, Cohen R, Denis F, Gaudelus J, Martinot A, Lery T, et al. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. Med Mal Infect 2016 May;46(3):117-122. [doi: 10.1016/j.medmal.2016.02.002] [Medline: 26987960]
58. Hall MA, Dugan E, Zheng B, Mishra AK. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? Milbank Q 2001 Dec;79(4):613-39, v [FREE Full text] [doi: 10.1111/1468-0009.00223] [Medline: 11789119]

59. Rooney T, Lynch A, Gray F, McDonald W, Walsh A. Growing up in a surveillance society: The changing spaces of childhood experience. University of New England. 2010. URL: https://rune-uat.une.edu.au/web/handle/1959.11/9236 [accessed 2022-10-24]
60. Marx G, Steeves V. From the beginning: children as subjects and agents of surveillance. SS 2010 Jun 17;7(3/4):192-230. [doi: 10.24908/ss.v7i3/4.4152]
61. Young climate activists demand action and inspire hope. UNICEF. URL: https://www.unicef.org/stories/young-climate-activists-demand-action-inspire-hope [accessed 2022-10-24]
62. Watts N, Amann M, Arnell N, Ayeb-Karlsson S, Belesova K, Boykoff M, et al. The 2019 report of The Lancet Countdown on health and climate change: ensuring that the health of a child born today is not defined by a changing climate. Lancet 2019 Nov 16;394(10211):1836-1878. [doi: 10.1016/S0140-6736(19)32596-6] [Medline: 31733928]
63. Zhang Y, Bi P, Hiller JE. Climate change and disability-adjusted life years. J Environ Health 2007 Oct;70(3):32-36. [Medline: 17941401]
64. "Lean ICT: Towards Digital Sobriety": Our new report on the environmental impact of ICT. The Shift Project. 2019 Mar 06. URL: https://theshiftproject.org/en/article/lean-ict-our-new-report/ [accessed 2022-10-24]
65. MacNeill AJ, McGain F, Sherman JD. Planetary health care: a framework for sustainable health systems. Lancet Planet Health 2021 Feb;5(2):e66-e68 [FREE Full text] [doi: 10.1016/S2542-5196(21)00005-X] [Medline: 33581064]
66. Jones N. How to stop data centres from gobbling up the world's electricity. Nature 2018 Sep 12;561(7722):163-166. [doi: 10.1038/d41586-018-06610-y] [Medline: 30209383]
67. Dayarathna M, Wen Y, Fan R. Data center energy consumption modeling: a survey. IEEE Commun. Surv. Tutorials 2016;18(1):732-794. [doi: 10.1109/comst.2015.2481183]
68. Strubell E, Ganesh A, McCallum A. Energy and Policy Considerations for Deep Learning in NLP. 2019 Presented at: 57th Annual Meeting of the Association for Computational Linguistics (ACL); July 2019; Florence, Italy URL: http://arxiv.org/abs/1906.02243 [doi: 10.18653/v1/p19-1355]
69. Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. Commun. ACM 2020 Nov 17;63(12):54-63. [doi: 10.1145/3381831]
70. Sustainability. National Institute for Health and Care Excellence (NICE). URL: https://www.nice.org.uk/about/who-we-are/sustainability [accessed 2022-10-24]
71. Mason E. Value Pluralism. In: Zalta EN, editor. Stanford Encyclopedia of Philosophy Archive. Stanford, CA: Stanford University; 2018.

Abbreviations

AI: artificial intelligence
CYP: children and young people
DT: digital twin
DTS: digital twin system
EMA: European Medicines Agency
FDA: Food and Drug Administration
IoT: internet of things

Edited by T Leung; submitted 18.05.22; peer-reviewed by E Hekler, B Nievas Soriano; comments to author 25.08.22; revised version received 11.09.22; accepted 11.10.22; published 31.10.22.

Please cite as:
Drummond D, Coulet A
Technical, Ethical, Legal, and Societal Challenges With Digital Twin Systems for the Management of Chronic Diseases in Children and Young People
J Med Internet Res 2022;24(10):e39698
URL: https://www.jmir.org/2022/10/e39698
doi:10.2196/39698
PMID:36315239

Original Paper

A Data-Driven Reference Standard for Adverse Drug Reaction (RS-ADR) Signal Assessment: Development and Validation

Suehyun Lee[1], PhD; Jeong Hoon Lee[2], PhD; Grace Juyun Kim[2], PhD; Jong-Yeup Kim[3], MD, PhD; Hyunah Shin[3], BS; Inseok Ko[3], MS; Seon Choe[2], MS, MD; Ju Han Kim[2], MD, PhD

[1]Department of Biomedical Informatics, College of Medicine, Konyang University, Daejeon, Republic of Korea

[2]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, Republic of Korea

[3]Healthcare Data Science Center, Konyang University Hospital, Daejeon, Republic of Korea

Corresponding Author:
Ju Han Kim, MD, PhD
Seoul National University Biomedical Informatics (SNUBI)
Division of Biomedical Informatics
Seoul National University College of Medicine
103 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 010 4636 2227
Email: juhan@snu.ac.kr

Abstract

Background: Pharmacovigilance using real-world data (RWD), such as multicenter electronic health records (EHRs), yields massively parallel adverse drug reaction (ADR) signals. However, proper validation of computationally detected ADR signals is not possible due to the lack of a reference standard for positive and negative associations.

Objective: This study aimed to develop a reference standard for ADR (RS-ADR) to streamline the systematic detection, assessment, and understanding of almost all drug-ADR associations suggested by RWD analyses.

Methods: We integrated well-known reference sets for drug-ADR pairs, including Side Effect Resource, Observational Medical Outcomes Partnership, and EU-ADR. We created a pharmacovigilance dictionary using controlled vocabularies and systematically annotated EHR data. Drug-ADR associations computed from MetaLAB and MetaNurse analyses of multicenter EHRs and extracted from the Food and Drug Administration Adverse Event Reporting System were integrated as “empirically determined” positive and negative reference sets by means of cross-validation between institutions.

Results: The RS-ADR consisted of 1344 drugs, 4485 ADRs, and 6,027,840 drug-ADR pairs with positive and negative consensus votes as pharmacovigilance reference sets. After the curation of the initial version of RS-ADR, novel ADR signals such as “famotidine–hepatic function abnormal” were detected and reasonably validated by RS-ADR. Although the validation of the entire reference standard is challenging, especially with this initial version, the reference standard will improve as more RWD participate in the consensus voting with advanced pharmacovigilance dictionaries and analytic algorithms. One can check if a drug-ADR pair has been reported by our web-based search interface for RS-ADRs.

Conclusions: RS-ADRs enriched with the pharmacovigilance dictionary, ADR knowledge, and real-world evidence from EHRs may streamline the systematic detection, evaluation, and causality assessment of computationally detected ADR signals.

(J Med Internet Res 2022;24(10):e35464) doi:10.2196/35464

KEYWORDS

adverse drug reaction; ADR; real-world data; RWD; real-world evidence; RWE; pharmacovigilance; PV; reference standard; pharmacology; drug reaction

Introduction

Theories

An increasing number of studies have reported serious postmarket adverse drug reactions (ADRs) that were not discovered in Phase III clinical trials. Clinical trials are inherently limited in reflecting real-world settings where patients with diverse demographics and comorbidities take a variety of concurrent medications [1]. Real-world factors such as off-label medication prescriptions and irregular drug intake increase the risk of missing ADRs in clinical trials. Clinical trials have difficulty in identifying ADRs occurring in the real-world environment, such as delayed ADRs and effects from long-term drug exposure [2]. ADR-related medical costs for morbidity and mortality in the United States have been reported to be greater than US $75 billion per year [3,4]. Hence, the importance of postmarket drug-safety surveillance cannot be overemphasized. Drug-safety surveillance plays a role in managing and preventing potential ADRs and involves a wide range of activities that includes an entire cycle of collecting, analyzing, and monitoring related to ADRs. ADR signals exist in many forms, such as clinical signs, symptoms, diseases, or deaths. Spontaneous reporting systems, collecting suspected postmarket ADRs with causality assessments [5], are inherently biased.

Prior Work

Computational methods for massively parallel detection of almost all drug-ADR interactions using real-world data (RWD), such as claims and multicenter electronic health records (EHRs), are emerging as relatively unbiased approaches [6-16]. However, validating massively detected ADR signals is challenging due to the lack of a “gold standard” or established reference set for all pairwise drug-ADR associations. In addition, determining a negative association is even more difficult than a positive one. Even the large, expert-curated reference standard provided by the major entities are disappointingly inadequate in correctly evaluating all computationally detected drug-ADR interactions. A reference standard involves a set of positive cases that are truly related to ADRs and negative controls that are highly unlikely to be associated. The reference standard should be formidable and have variety with multiple drugs and ADRs to ensure generalizability [17].

Coloma et al [10] developed a reference standard with 44 positive and 50 negative associations. The Observational Medical Outcomes Partnership (OMOP) presented a comprehensive compilation of 165 positive and 234 negative outcomes from their resources [18]. The EU-ADR presented 10 types of events associated with drug use, including 44 positive and 50 negative controls, based on a literature review [10]. Recently, Observational Health Data Sciences and Informatics published a knowledge base of 1000 drugs and 100 health outcomes of interest [19]. The Observational Health Data Sciences and Informatics group developed and tested the accuracy of an automated reference set to reduce manual curations [20]. Considering that previous studies [18-21] have relied mainly on literature and spontaneous reports, the coordination of evidence from different data sources is needed.

In silico ADR detection using RWD is much faster than reference standard development relying on expert curations. RWD analysis can potentially provide a reference standard for ADR signal evaluation. A systematic application of controlled vocabularies with rich semantics is essential for in silico pharmacovigilance (PV) using RWD. The controlled vocabulary–based ADR signal dictionary (CVAD) integrated controlled vocabularies with EHR data to improve PV [22]. The development of CVAD was motivated by previous research on massively parallel ADR signal detection algorithms using laboratory results and standard nursing statements, MetaLAB and MetaNurse [23]. Given the limited numbers of positive and negative reference sets, the correct validation of positive and negative drug-ADR associations among 101 precautionary drugs by thousands of ADR signals is challenging. A comprehensive reference standard is required for drug-ADR pairs, equipped with standard vocabulary annotations, in the emerging era of RWD and real-world evidence (RWE).

For prevention and management in PV, a strategy for integrating multiple data sources is preferred. Wei et al [24,25] combined RxNorm, Side Effect Resource (SIDER), MedlinePlus, and Wikipedia to compose a medication indication resource (MEDI). Gottesman et al [26] developed the Electronic Medical Records and Genomics network that advanced clinical informatics, genome science, and community consultation as a first step toward incorporating genomic information into routine health care delivery. Additionally, national-level projects are being carried out in several countries, or related research authorized, due to the need for a data-driven approach.

Goal of This Study

A key challenge in drug-safety surveillance, regardless of data source, is that publicly available, reliable, and sufficiently large reference standards are needed. Although no definitive reference standard contains a complete set of ADRs, we intended to aggregate information from multiple data sources to constitute a set. In this study, we developed a reference standard for ADR (RS-ADR) for the comprehensive, efficient, and pragmatic evaluation of computationally detected massive ADR signals from RWD. RS-ADR integrates EHR term–related standard ADR terminologies, including those from the Medical Dictionary for Regulatory Activities (MedDRA) preferred terms (PTs), WHO Adverse Reactions Terminology (WHO-ART), Logical Observation Identifiers Names and Codes (LOINC), and International Classification of Diseases 10th Revision (ICD-10). We created the RS-ADR by aggregating massively parallel results of RWD and cross-validations for the positive and negative cases extracted from a multitude of health care organizations. Other PV resources, including OMOP and EU-ADR reference standards, the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) [27], and SIDER 4.1, were used as reference sets and augmented with controlled PV vocabularies to improve systematic causality assessments of drug-ADR associations. We tried to analyze and compare previously published reference sets, significantly increasing the number of cases, and developed RS-ADR by focusing on terminology standardization.

Methods

Development of Reference Sets for ADR Signal Evaluation

Given the lack of a "gold standard" to evaluate true and false ADR signals detected from PV studies, many researchers have attempted to compose ad hoc "gold standard" sets. Table 1 summarizes the different characteristics of the proposed "gold standard" sets by data source, number of drugs and ADRs, numbers of drug-ADR pairs, true positive and negative cases, controlled vocabularies, and evidence. The main objective of these studies in creating these reference standards was to evaluate the performance of the proposed algorithms. ADR signals were mainly defined by laboratory test results and clinical events, such as symptoms.

Table 1. Reference sets created and used by pharmacovigilance methodological studies.

Reference set	Data source	Drugs, n	ADRs[a]				Vocabulary
			ADRs, n	Drug-ADR pairs, n	Positive cases, n	Negative cases, n	
RS-ADR[b]	Laboratory test and event (symptom)	1344	4485	6,027,840	141,729	2330	ATC[c], MedDRA[d], WHO-ART[e], LOINC[f], and ICD[g]
Harpaz et al [13], 2012	Event (symptom)	44	38	137	62	75	RxNorm, and MedDRA
Yoon et al [14], 2012	Laboratory test	10	51	510	—[h]	—	None
Liu et al [15], 2013	Laboratory test	9	42	378	—	—	None
LePendu et al [16], 2013	Event (symptom)	78	12	193	28	165	MedDRA
Alvarez et al [28], 2010	Event (symptom)	267	—	—	532	—	MedDRA
Hochberg et al [29], 2009	Event (symptom)	35	—	—	6207	—	MedDRA
Ryan et al [30], 2012 (OMOP[i] version 1)	Event (symptom)	10	9	90	9	44	None
Ryan et al [18], 2013 (OMOP version 2)	Event (symptom)	191	4	398	165	234	None
Coloma et al [10], 2013 (EU-ADR)	Event (symptom)	66	10	94	43	50	None
Boyce et al [19], 2014 (OHDSI[j] knowledge base)	Event (symptom)	1000	100	100,000	—	—	ATC, RxNorm, and ICD

[a]ADR: adverse drug reaction.

[b]RS-ADR: reference standard for adverse drug reaction.

[c]ATC: Anatomical Therapeutic Chemical.

[d]MedDRA: Medical Dictionary for Regulatory Activities.

[e]WHO-ART: World Health Organization Adverse Reactions Terminology.

[f]LOINC: Logical Observation Identifiers Names and Codes.

[g]ICD: International Classification of Diseases.

[h]Not available.

[i]OMOP: Observational Medical Outcomes Partnership.

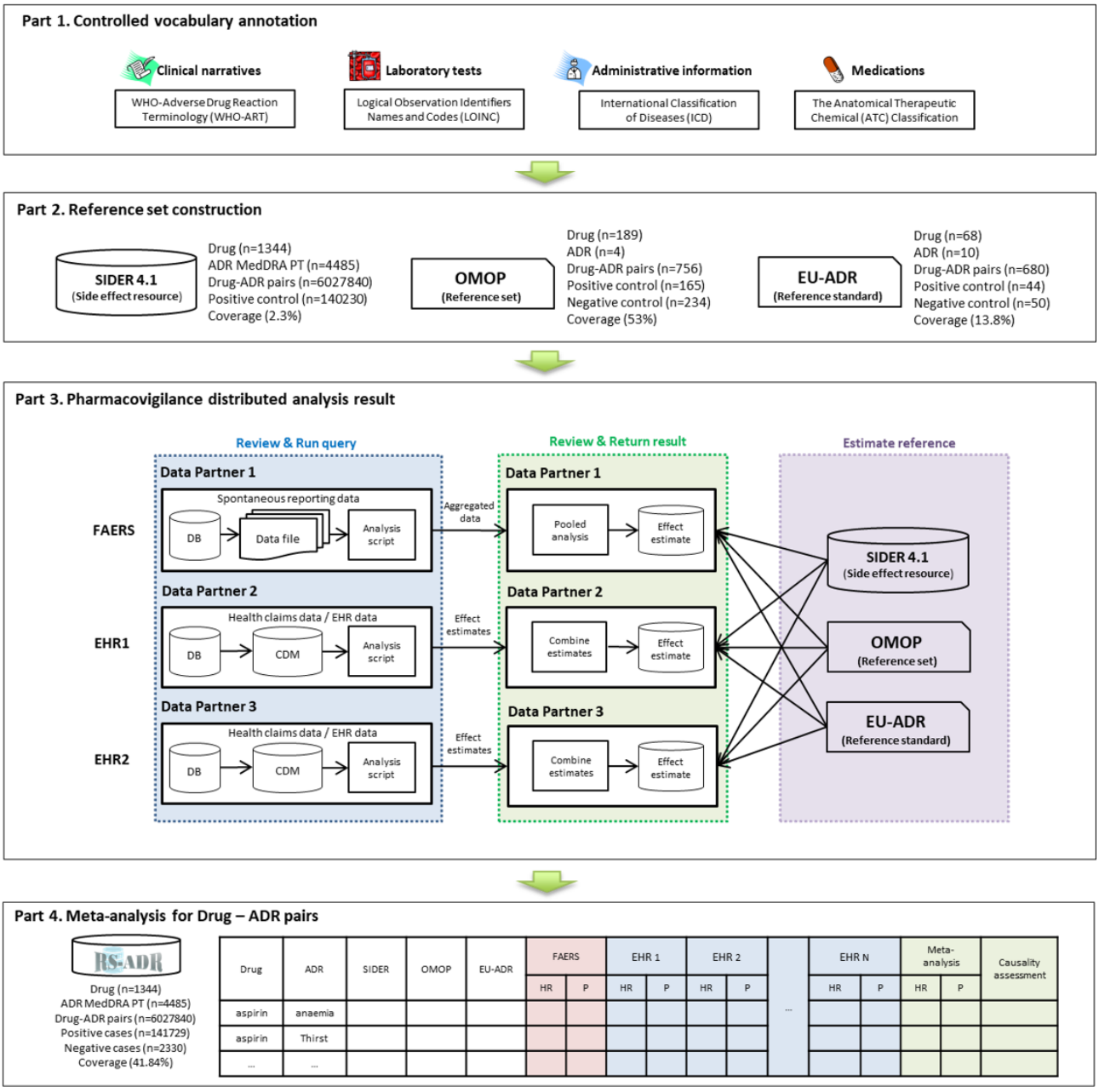
[j]OHDSI: Observational Health Data Sciences and Informatics.

The practical databases used in the study for constructing RS-ADRs were SIDER 4.1, OMOP, and EU-ADR. SIDER 4.1 contains the numbers of drugs, ADRs, drug-ADR pairs, and drug frequency entries from various references [31]. In addition, there are various databases (eg, Sentinel and the National Patient-Centered Clinical Research Network), but OMOP and EU-ADR are the most used in all fields of PV and provide an actual reference set. The researchers manually reviewed related references and finally selected the databases after being confirmed by clinicians. The OMOP database derived data from private contractors in the United States and EU-ADR derived data from European nationwide registries. Both were used for the identification of well-known drug associations and previously unknown signals [32].

A reference standard is essential for the evaluation of analysis results and systematic accumulation of evidence from comprehensive PV studies [13-31]. Therefore, we first created

a reference standard based on the OMOP and EU-ADR projects. In all, 4 steps were used in the construction of the RS-ADR: (1) controlled vocabulary annotation, (2) reference set construction, (3) distributed analysis results, and (4) meta-analysis for drug-ADR pairs (Figure 1). The role of each part is elaborated in the following sections.

Figure 1. Flowchart of the construction of the RS-ADR, which uses electronic health record data (clinical narrative, laboratory tests, and disease classification). ADR: adverse drug reaction; CDM: common data model; DB: database; EHR: electronic health record; FAERS: Food and Drug Administration Adverse Event Reporting System; HR: hazard ratio; MedDRA: Medical Dictionary for Regulatory Activities; OMOP: Observational Medical Outcomes Partnership; PT: preferred term; RS-ADR: reference standard for adverse drug reaction.



Part 1: Controlled Vocabulary Annotation

A comprehensive annotation of controlled vocabularies that encompass disease classifications, laboratory tests, medications, and clinical narratives enables extensive EHR data exploration. Laboratory results have been used the most frequently for ADR signal detection in many studies. CVAD facilitates the use of a variety of data sources to detect ADR signals [22]. Clinical narratives such as International Council for Nursing Practice–based standard nursing statements (SNSs) of Seoul National University Hospital (SNUH) were mapped to WHO-ART, laboratory test results from SNUH or Ajou University Hospital were mapped to LOINC, administrative terms were mapped to ICD-10, and medications were mapped to ATC classifications. The mapping schemes involving narrative, laboratory, or administrative terms have been described in detail elsewhere [22].

Part 2: Reference Set Construction

OMOP also provides a reference set, which is composed of 165 positive and 234 negative drug-ADR signal pairs, covering 53% of the 756 (189 × 4) pairs between 189 drugs and 4 ADRs [15].

The reference set of the EU-ADR project covers 68 drugs and 10 ADRs with 44 positive and 50 negative drug-ADR signal pairs, covering 13.8% [16]. The reference set prepared by SIDER 4.1 includes 140,230 positive pairs in MedDRA PTs for 1344 drugs and 4485 side effects without providing negative controls [31]. The coverage of SIDER 4.1 for drug-ADR pairs was 2.3%.

We mapped the 4 ADRs of OMOP and the 10 ADRs of EU-ADR to 4485 MedDRA terms in SIDER 4.1 using MedDRA synonyms (Unified Medical Language System Concept ID). We created a reference standard matrix for 1344 drugs and 4485 ADRs returning 6,027,840 drug-ADR pairs. The value of each cell of the reference matrix was filled with 0 for negative controls, 1 for positive controls, and 2 for unknowns. Negative controls were those known to not cause the outcomes, using case reports, case series, or observational evidence in OMOP and EU-ADR. Positive controls were extracted from the product labels in the US FDA "Black Box Warning" section in SIDER 4.1, OMOP, and EU-ADR (Figure 1).

Part 3: Distributed PV Analysis Results

We benchmarked the analysis data from various institutions for PV [18] and integrated results from various resources such as spontaneous reports (ie, FAERS data), claims, and EHR data for developing RS-ADR. FAERS data from January 2012 to December 2018 were analyzed [27]. We performed the MetaNurse and MetaLAB analyses for the entire EHR data sets of 2 hospitals (SNUH and Konyang University Hospital) for SNSs and laboratory results using an advanced subject-sampling strategy for managing drugs, laboratory results, and SNSs. The detected ADR signals from the 2 EHR data sets were validated against SIDER 4.1 using 11,817 and 76,457 drug-ADR pairs, respectively [23]. We explored the relationship between drug-ADR pairs using spontaneous reports and EHR data. Table 2 shows the consensus template of our validation efforts for the "fluconazole-hypokalemia" association detected by the algorithms. Previous studies without annotated, controlled vocabularies experienced difficulty in evaluating their study results [33].

Table 2. Example of RS-ADR[a] output for the association between "fluconazole" and "hypokalemia."

Output, name	Example
Drug	
Drug	Fluconazole
ATC[b] code	J02AC01
ADR[c]	
MedDRA[d] PT[e]	Hypokalemia
System organ class	Metabolism and nutrition disorders
Part 1	
Clinical narrative	
WHO-ART[f]	Hypokalemia
SNS[g] terms at SNUH[h]	Serum potassium levels under normal | Hypokalemia
ICNP[i]	"mg/dL," "not balanced," and "fluid volume"
Laboratory results	
LOINC[j] ID	2823_3
LOINC common name	Potassium (moles/volume) in serum or plasma
SNUH laboratory test code	L3044
SNUH laboratory test name	Potassium (serum)
AJUH[k] laboratory test code	35
AJUH laboratory test name	Potassium
Disease classification	
ICD[l] code	E87.6
ICD name	Hypokalemia
Part 2	
Evidence source (0=negative control, 1=positive control, and 2=unknown)	
FDA[m] product label**:** SIDER[n]	1
FDA product label and literature: OMOP[o]	2
FDA product label, literature, spontaneous data, and mechanism of action: EU-ADR	2
Part 3	
Data partner: SNUH (EHR[p]-based MetaNurse)	
Hazard ratio	1.47
P value	<.001
Data partner: SNUH (EHR-based MetaLAB)	
Odds ratio	3.04
P value	<.001
Data partner: KYUH[q] (EHR-based MetaLAB)	
Odds ratio	1.58
P value	<.001
Data partner: FAERS[r]	
Reporting odds ratio	1.83
P value	<.001

Output, name	Example
Data partner (N)	
Odds ratio	—[s]
P value	—
Part 4	
Meta-analysis	
Odds ratio (95% CI)	1.69 (1.60-1.79)
Causality assessment	possible

[a]RS-ADR: reference standard for adverse drug reaction.
[b]ATC: Anatomical Therapeutic Chemical.
[c]ADR: adverse drug reaction.
[d]MedDRA: Medical Dictionary for Regulatory Activities.
[e]PT: preferred term.
[f]WHO-ART: World Health Organization Adverse Reactions Terminology.
[g]SNS: standard nursing statement.
[h]SNUH: Seoul National University Hospital.
[i]ICNP: International Council for Nursing Practice.
[j]LOINC: Logical Observation Identifiers Names and Codes.
[k]AJUH: Ajou University Hospital.
[l]ICD: International Classification of Diseases.
[m]FDA: Food and Drug Administration.
[n]SIDER: Side Effect Resource.
[o]OMOP: Observational Medical Outcomes Partnership.
[p]EHR: electronic health record.
[q]KYUH: Konyang University Hospital.
[r]FAERS: Food and Drug Administration Adverse Event Reporting System.
[s]Not available.

Part 4: Meta-analysis for Drug-ADR Pairs

We evaluated the drug-ADR pairs of the MetaLAB and MetaNurse analyses from multiple EHRs and compared with FAERS for causality assessments as follows: certain, probable/likely, possible, unlikely, or conditional/unclassified [10]. We applied a random-effects model for the meta-analysis of many results to manage the heterogeneous data characteristics of spontaneous reports and EHRs. To assess causality, we carried out expert reviews by having the experts refer to SIDER 4.1 and other existing references. Subsequently, PV-distributed analysis results generated by various health care organizations were collected for a causality assessment of each drug-ADR pair. With an increasing number of data partners providing study results, the causality assessment of each drug-ADR pair can be improved.

Ethical Considerations

This study was approved by the Institutional Review Board of Konyang University Hospital (IRB no 2019-08-018).

Results

RS-ADR Statistics

The RS-ADR contained 1344 drugs and 4485 ADRs in terms of MedDRA PTs (Tables 3 and 4). The number of controlled vocabularies mapped to MedDRA PTs was for 1130 clinical narratives, 942 laboratory results, and 83 disease classifications. For positive controls, we found 140,230 drug-ADR pairs from SIDER 4.1, 1556 from OMOP, and 421 from the EU-ADR databases. The negative controls were 2801 and 349 drug-ADR pairs from OMOP and EU-ADR, respectively. ADRs were examined according to a variety of MedDRA system organ classes (SOCs) for clinical narratives, laboratory results, and disease classifications, covering 25, 23, and 16 of the 26 MedDRA SOCs, respectively (Multimedia Appendix 1). Although previous ADR studies predominantly analyzed laboratory results, we browsed 1762 integrative ADRs (ie, the intersection of clinical narrative, laboratory tests, and disease classification) with RS-ADR.

Table 3. RS-ADR[a] statistics.

Statistic	Value, n
Drugs	1344
ADRs[b] (MedDRA[c] preferred term)	
Total	4485
Clinical narrative	1130
Laboratory tests	942
Disease classification	83
Not mapped	2723
Drug-ADR pairs (number of drugs × number of ADRs)	6,027,840

[a]RS-ADR: reference standard for adverse drug reaction.

[b]ADR: adverse drug reaction.

[c]MedDRA: Medical Dictionary for Regulatory Activities.

Table 4. RS-ADR[a] statistics in comparison with other reference sets.

Statistic	SIDER[b]	OMOP[c]	EU-ADR
Positive controls, n	140,230	1556	421
Negative controls, n	—[d]	2801	349
Unknown drug-ADR[e] pairs, n	5,887,610	6,023,483	6,027,070

[a]RS-ADR: reference standard for adverse drug reaction.

[b]SIDER: Side Effect Resource.

[c]OMOP: Observational Medical Outcomes Partnership.

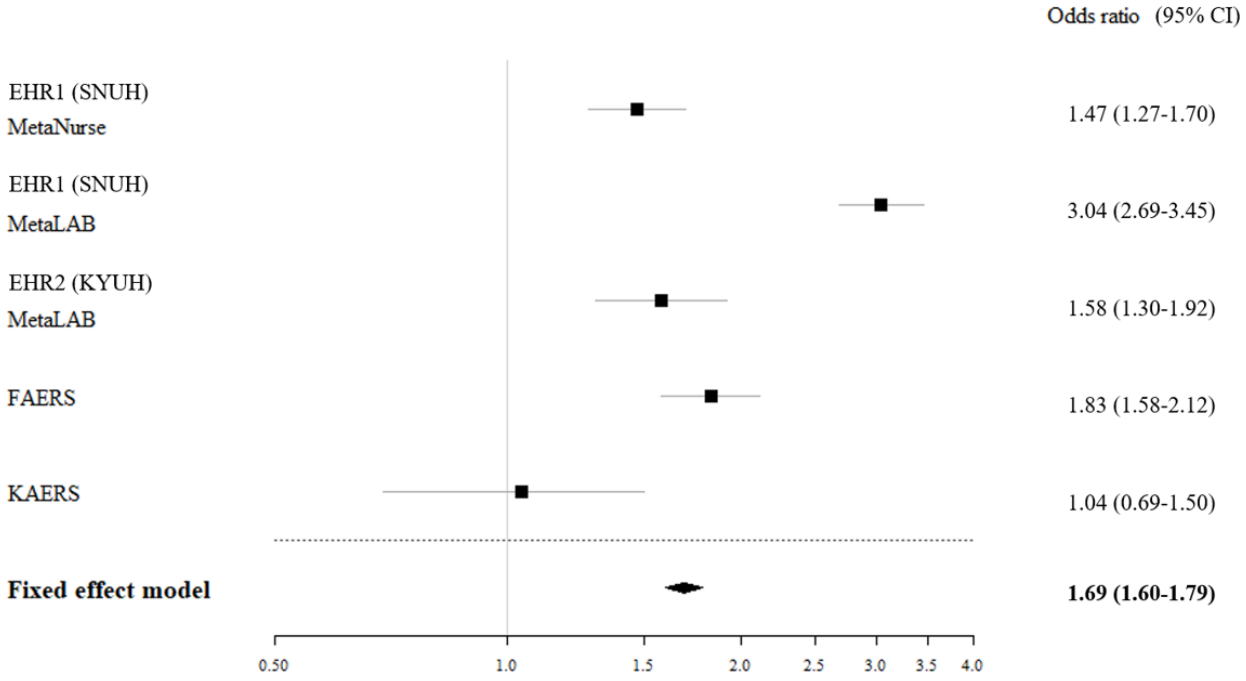
[d]Not available.

[e]ADR: adverse drug reaction.

An Example Application of RS-ADR

The process from part 1 to 4 for RS-ADR construction is briefly summarized as follows: first, the drugs and ADRs to be targeted; in part 1, term code confirmation; in part 2, the identification of contents described in the existing reference set; in part 3, analysis by data source; and in part 4, causality evaluation through meta-analysis. Table 2 shows a query result from RS-ADR for the association between "fluconazole" and "hypokalemia," which explains the progress in stages from part 1 to part 4 in order. Part 1 consisted of 3 components: clinical narratives, laboratory results, and administrative data. Clinical narratives were annotated with WHO-ART "hypokalemia"; SNS "serum potassium levels less than normal"; and International Council for Nursing Practice "mg/dL," "not balanced," and "fluid volume" standard terms. Laboratory results were mapped to up to 6 tests, including LOINC "potassium (moles/volume) in serum or plasma." The RS-ADR also indicated the direction of the test result to be higher or lower than the normal range. The administrative term mapped to ADR hypokalemia was the ICD-10 E87.6 code. Part 2 presented the evidence source of the drug-ADR association with positive controls, negative controls, and unknown evidence. Evidence sources could be FDA product labels, literature, spontaneous reports, and mechanisms of action. Part 3 designated the partner health care organizations where the ADR analysis data were collected. MetaLAB and MetaNurse analyses were included [23]. Finally, part 4 described how the causality between drug-ADR occurrence was assessed. A meta-analysis of the association between "fluconazole" and "hypokalemia" showed an odds ratio of 1.69 (95% CI, 1.60–1.79). In all, 2 EHRs and 2 spontaneous reporting data sets show the scalability and availability of the RS-ADR (Figure 2). The usability of RS-ADR can be enhanced by adding drug-ADR pairs using RWD analysis. The association between "fluconazole" and "hypokalemia" was assessed according to the WHO–Uppsala Monitoring Centre causality categories as "possible," as this category included the criteria "event or laboratory test abnormality" (Table 2) [33].

Figure 2. Example of the RS-ADR (part 3) for evaluating the association between the drug "fluconazole" and "hypokalemia" by using electronic health records (EHRs) from 2 hospitals (Seoul National University Hospital [SNUH] and Konyang University Hospital [KYUH]) and Food and Drug Administration Adverse Event Reporting System (FAERS) data. RS-ADR: reference standard for adverse drug reaction.



Improving Reference Standards Using RWE

Table 5 shows 4 drug-ADR pairs that were previously unknown in SIDER 4.1, OMOP, and EU-ADR. In this regard, we found that 2 of the drug-ADR pairs were added to Korean FDA ADR labels [34], which signals that they might have been determined as false positives. For example, famotidine was used in gastrointestinal conditions related to acid secretion (eg, gastric ulcers) and gastroesophageal reflux disease [35]. The novel "famotidine–hepatic function abnormal" pair discovered by RS-ADR was successfully validated by 2 institutional EHRs and by US FAERS [35]. Moreover, according to the Micromedex [36] database and a study by Gupta et al [37], we found that the famotidine–hepatic function abnormal pair had been documented as a possible ADR. The RWD/RWE perspective suggests that the novel finding may indeed indicate a true positive supported by multi-institutional cross-validations. We performed the same analysis for clozapine and diclofenac and found reasonable support (with reservations) for the potential drug-ADR pairs "clozapine–hepatic function abnormal," "diclofenac-angioedema," and "diclofenac–face edema" (Table 5).

Table 5. RS-ADR[a] evidence of how significant the drug-ADR[b] pairs are using the EHR[c] data of 2 hospitals (Seoul National University Hospital [SNUH] and Konyang University Hospital [KYUH]) and Food and Drug Administration Adverse Event Reporting System (FAERS) data.

Drug	ADR	RS-ADR								Reference
		SNUH				KYUH		FAERS		
		EHR-based MetaNurse		EHR-based MetaLAB		EHR-based MetaLAB				
		HR[d]	*P* value	OR[e]	*P* value	OR	*P* value	OR	*P* value	
Famotidine	Hepatic function abnormal	1.79	<.001	2.19	.003	1.11	.008	3.97	<.001	Gupta et al [37], 2009
Clozapine	Hepatic function abnormal	0.55	.04	1.38	.01	—[f]	—	1.02	.85	Wu Chou et al [38], 2014
Diclofenac	Angioedema	0.96	.47	—	—	—	—	5.13	<.001	Pise and Padwal [39], 2015
Diclofenac	Face edema	2.38	.20	—	—	—	—	1.95	<.001	Jha et al [40], 2015

[a]RS-ADR: reference standard for adverse drug reaction.

[b]ADR: adverse drug reaction.

[c]EHR: electronic health record.

[d]HR: hazard ratio.

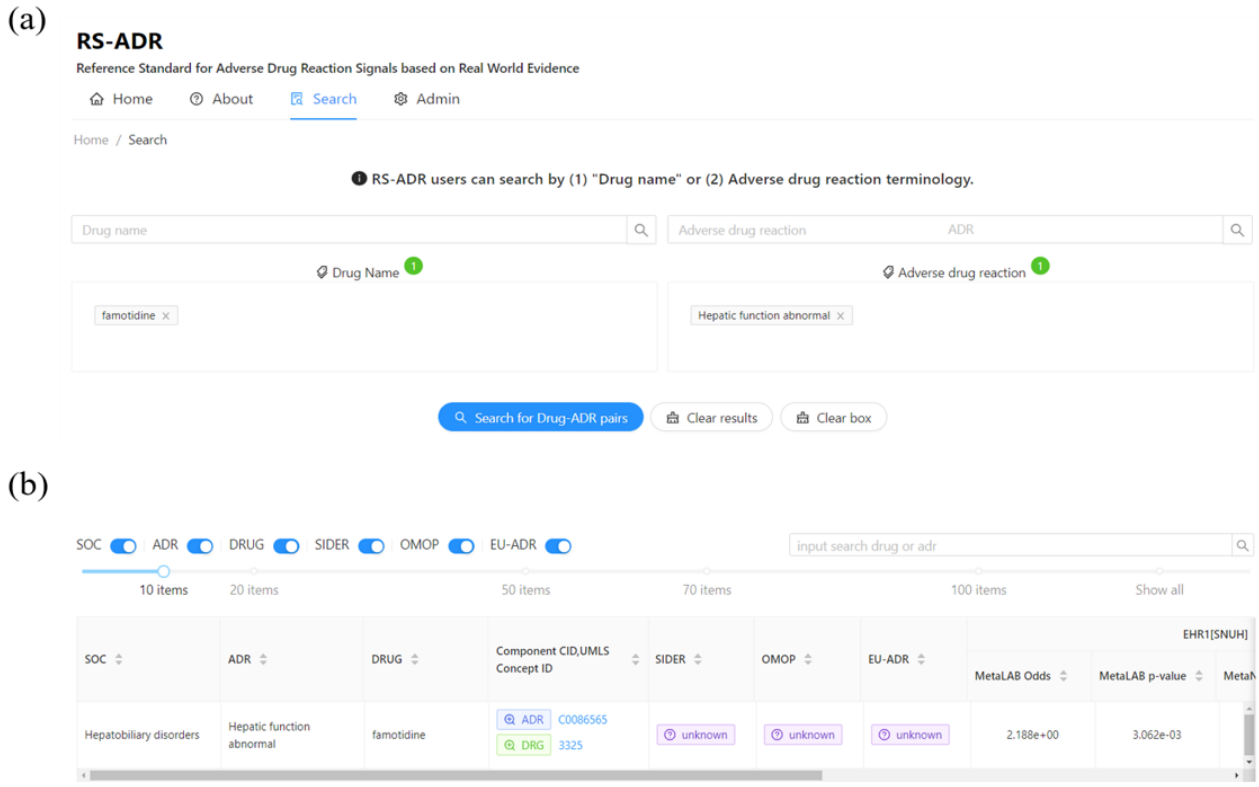
[e]OD: odds ratio.

[f]Not available.

Web-Based RS-ADR Explorer

To provide a semantically enriched ADR dictionary for postmarket drug safety research and enable multicenter EHR-based extensive ADR signal evaluation, we developed a web-based search interface for RS-ADR to explore drug-ADR associations [41] (Figure 3). Figure 3 shows the drug-ADR search functions and the results of a "famotidine–hepatic function abnormal" query. Users can search for interesting drug-ADR pairs in combination; each search function adds similar words using drop-down menus. A button clears the drug-ADR combinations and results to facilitate searching. Search results appear in the order of SOC, ADR, drug, additional information (component identification for drug and Unified Medical Language System concept ID for ADR), comparison of reference standards (SIDER, OMOP, and EU-ADR), and each result of the EHR and FAERS (odds ratio and *P* value). Parts 3 and 4 of the RS-ADR have a structure that allows researchers to add and update their results to improve the RS-ADR.

Figure 3. User interface for the RS-ADR for exploring the drug-ADR relationship. (A) Drug-ADR search; (B) Example of RS-ADR query: association between "famotidine" and "hepatic function abnormal." ADR: adverse drug reaction; CID: component identification; OMOP: Observational Medical Outcomes Partnership; RS-ADR: reference standard for adverse drug reaction; SIDER: Side Effect Resource; SOC: system organ class; UMLS: Unified Medical Language System.



Discussion

Principal Findings

In this study, we demonstrated the possibility of creating an RWD-based RS-ADR. We integrated various standard vocabularies to facilitate the use of different institutional EHR databases along with other PV resources, such as SIDER 4.1, OMOP, and EU-ADR. Integrative analysis of heterogeneous real-world clinical information requires a standard vocabulary to correctly interpret study results.

The reference sets of OMOP and EU-ADR [15,16] are difficult to apply directly in PV research, because they only provide information about the relationships between the selected drugs and ADRs. To use these reference sets, each observational database should be reconstructed and annotated using controlled vocabularies by the researchers. The RS-ADR approach facilitates the accumulation of RWD-driven evidence extracted from various sources, including many EHRs and claims databases. The scope of detectable ADRs was widely expanded by RS-ADR using FDA structured product labels and low ADR concept levels (eg, MedDRA PTs). A low ADR concept level is most commonly used in the standard terminology system to explain detailed symptoms such as MedDRA PTs. RS-ADR complements this limitation by establishing a reference standard using 1344 drugs and 4485 ADRs. The RS-ADR approach used in this study is not as biased toward positive findings as other PV resources but is balanced between positive and negative drug-ADR associations due to its unbiased computational approach. Multimedia Appendix 1 shows the distribution of MedDRA PT–annotated ADRs detected using clinical narratives, laboratory results, and administrative terms grouped by SOCs. The SOCs "infections and infestations," "psychiatric disorders," and "eye disorders" exhibit many ADRs that are difficult to detect from laboratory results only and require clinical narratives, nursing statements, and administrative terms in the RS-ADR. The ADRs in "musculoskeletal and connective tissue

disorders" and "ear and labyrinth disorders" SOCs could only be found using clinical narratives.

Limitations

Our study has some limitations. SIDER 4.1 provides inadequate information about postmarket ADRs as it comprises public documents and package inserts. The 4 ADRs of the OMOP and 10 ADRs of the EU-ADR project may emphasize ADRs of more frequently or chronically used drugs, which are also clinically important. The use of integrative ADR references such as SIDER 4.1, OMOP, and EU-ADR in the RS-ADR complements the limitations of each resource. Although the RS-ADR went through interevaluator agreement, expert evaluation was substantially limited, and continuous review and updates are required. When integrated with multicenter and multinational data, RS-ADR becomes a meaningful RWE-based reference standard for evaluating ADR signals. Underlying the use of a reference standard for method evaluation is the assumption that negative controls are exchangeable with positive controls [10,18]. Adding drug-ADR pairs from various studies to the RS-ADR can increase its evidence base and is a topic of future research. In addition, considering the continuous RS-ADR update, it is planned to manage the analysis of new drugs and whether to discontinue the use of existing drugs. For national use, since the Korean Ministry of Food and Drug Safety is conducting related research (eg, multicenter analysis using common data model–based EHR, analyzing each drug-ADR pair), our team will contemplate various utilizations of RS-ADR for collecting and evaluating the research. Conversely, recent attempts to study ADRs related to herbal medicines have steadily increased [42-44], and we consider that it may be possible to apply RS-ADR construction to the field of herbal medicine in the future.

Conclusions

RS-ADR enriched with the PV dictionary, knowledge, and RWE can streamline the systematic detection, evaluation, and causality assessments of computationally detected ADR signals. Through RS-ADR, evidence related to ADRs can be prepared as much as possible before the clinical evaluation stage, and we could identify more cases based on actual medical center data—RWD. In addition, since we considered the standardization of terms for drugs and ADRs, it is highly useful when adding medical center or other resources in the future. It is applicable not only to ADR studies but also to a variety of health outcomes and health care database utilization studies.

Acknowledgments

The authors would like to acknowledge the National Research Foundation of Korea (NRF) and Ministry of Food and Drug Safety. This study was funded by the Ministry of Education (NRF-2021R1I1A3044287).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of adverse drug reactions in MedDRA preferred terms of the RS-ADR for each system organ class. ADR: adverse drug reaction; MedDRA: Medical Dictionary for Regulatory Activities; RS-ADR: reference standard for adverse drug reaction.
[PNG File , 55 KB - jmir_v24i10e35464_app1.png]

References

1. Holland J. Fixing a broken drug development process. J Commer Biotechnol 2013;19(1):5-6 [FREE Full text] [doi: 10.5912/jcb588]
2. Yadav S. Status of adverse drug reaction monitoring and pharmacovigilance in selected countries. Indian J Pharmacol 2008 Feb;40(Suppl 1):S4-S9 [FREE Full text] [Medline: 21369413]
3. Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, et al. A time-indexed reference standard of adverse drug reactions. Sci Data 2014 Nov 11;1:140043 [FREE Full text] [doi: 10.1038/sdata.2014.43] [Medline: 25632348]
4. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. J Gen Intern Med 2003 Jan;18(1):57-60 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20130.x] [Medline: 12534765]
5. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. BMC Bioinformatics 2010 Oct 28;11(Suppl 9):S7 [FREE Full text] [doi: 10.1186/1471-2105-11-S9-S7] [Medline: 21044365]
6. Ghosh P, Dewanji A. Effect of reporting bias in the analysis of spontaneous reporting data. Pharm Stat 2015;14(1):20-25. [doi: 10.1002/pst.1657] [Medline: 25376637]
7. Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. J Biomed Inform 2003;36(1-2):131-143 [FREE Full text] [doi: 10.1016/j.jbi.2003.08.003] [Medline: 14552854]
8. Penz JFE, Wilcox AB, Hurdle JF. Automated identification of adverse events related to central venous catheters. J Biomed Inform 2007 Apr;40(2):174-182 [FREE Full text] [doi: 10.1016/j.jbi.2006.06.003] [Medline: 16901760]

9. Honigman B, Lee J, Rothschild J, Light P, Pulling RM, Yu T, et al. Using computerized data to identify adverse drug events in outpatients. J Am Med Inform Assoc 2001 May 01;8(3):254-266 [FREE Full text] [doi: 10.1136/jamia.2001.0080254] [Medline: 11320070]
10. Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Saf 2013 Jan;36(1):13-23. [doi: 10.1007/s40264-012-0002-x] [Medline: 23315292]
11. Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc 2010 Nov 01;17(6):652-662 [FREE Full text] [doi: 10.1136/jamia.2009.002477] [Medline: 20962127]
12. Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. Ann Intern Med 2020 Jun 02;172(11 Suppl):S79-S84 [FREE Full text] [doi: 10.7326/M19-0873] [Medline: 32479175]
13. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther 2012 Jun;91(6):1010-1021 [FREE Full text] [doi: 10.1038/clpt.2012.50] [Medline: 22549283]
14. Yoon D, Park MY, Choi NK, Park BJ, Kim JH, Park RW. Detection of adverse drug reaction signals using an electronic health records database: Comparison of the Laboratory Extreme Abnormality Ratio (CLEAR) algorithm. Clin Pharmacol Ther 2012 Mar;91(3):467-474. [doi: 10.1038/clpt.2011.248] [Medline: 22237257]
15. Liu M, McPeek Hinz ER, Matheny ME, Denny JC, Schildcrout JS, Miller RA, et al. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. J Am Med Inform Assoc 2013 May 01;20(3):420-426 [FREE Full text] [doi: 10.1136/amiajnl-2012-001119] [Medline: 23161894]
16. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther 2013 Jun;93(6):547-555. [doi: 10.1038/clpt.2013.47] [Medline: 23571773]
17. Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, et al. A time-indexed reference standard of adverse drug reactions. Sci Data 2014 Nov 11;1(1):140043 [FREE Full text] [doi: 10.1038/sdata.2014.43] [Medline: 25632348]
18. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf 2013 Oct;36(Suppl 1):S33-S47. [doi: 10.1007/s40264-013-0097-8] [Medline: 24166222]
19. Boyce RD, Ryan PB, Norén GN, Schuemie MJ, Reich C, Duke J, et al. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Saf 2014 Aug;37(8):557-567 [FREE Full text] [doi: 10.1007/s40264-014-0189-0] [Medline: 24985530]
20. Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. J Biomed Inform 2017 Feb;66:72-81 [FREE Full text] [doi: 10.1016/j.jbi.2016.12.005] [Medline: 27993747]
21. Hauben M, Aronson JK, Ferner RE. Evidence of misclassification of drug-event associations classified as gold standard 'negative controls' by the Observational Medical Outcomes Partnership (OMOP). Drug Saf 2016 May;39(5):421-432. [doi: 10.1007/s40264-016-0392-2] [Medline: 26879560]
22. Lee S, Han J, Park RW, Kim GJ, Rim JH, Cho J, et al. Development of a controlled vocabulary-based adverse drug reaction signal dictionary for multicenter electronic health record-based pharmacovigilance. Drug Saf 2019 May;42(5):657-670. [doi: 10.1007/s40264-018-0767-7] [Medline: 30649749]
23. Lee S, Choi J, Kim H, Kim GJ, Lee KH, Park CH, et al. Standard-based comprehensive detection of adverse drug reaction signals from nursing statements and laboratory results in electronic health records. J Am Med Inform Assoc 2017 Jul 01;24(4):697-708. [doi: 10.1093/jamia/ocw168] [Medline: 28087585]
24. Wei WQ, Cronin RM, Xu H, Lasko TA, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. J Am Med Inform Assoc 2013 Apr 10;20(5):954-961 [FREE Full text] [doi: 10.1136/amiajnl-2012-001431] [Medline: 23576672]
25. Wei WQ, Mosley JD, Bastarache L, Denny JC. Validation and enhancement of a computable medication indication resource (MEDI) using a large practice-based dataset. AMIA Annu Symp Proc 2013 Nov 16;2013:1448-1456 [FREE Full text] [Medline: 24551419]
26. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
27. U.S. Food and Drug Administration. URL: https://www.fda.gov/drugs/guidance-compliance-regulatory-information [accessed 2018-05-02]
28. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. Drug Saf 2010 Jun 01;33(6):475-487. [doi: 10.2165/11534410-000000000-00000] [Medline: 20486730]

29. Hochberg AM, Hauben M, Pearson RK, O'Hara DJ, Reisinger SJ, Goldsmith DI, et al. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. Drug Saf 2009 Jun;32(6):509-525. [doi: 10.2165/00002018-200932060-00007] [Medline: 19459718]
30. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med 2012 Dec 30;31(30):4401-4415. [doi: 10.1002/sim.5620] [Medline: 23015364]
31. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 2010 Jan 19;6(1):343 [FREE Full text] [doi: 10.1038/msb.2009.98] [Medline: 20087340]
32. Lopes P, Nunes T, Campos D, Furlong LI, Bauer-Mehren A, Sanz F, et al. Gathering and exploring scientific knowledge in pharmacovigilance. PLoS One 2013 Dec 11;8(12):e83016 [FREE Full text] [doi: 10.1371/journal.pone.0083016] [Medline: 24349421]
33. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet 2000 Oct 07;356(9237):1255-1259. [doi: 10.1016/S0140-6736(00)02799-9] [Medline: 11072960]
34. Yoo S. Notification of change of drug product approval information (diclofenac beta-diaminoethanol ingredient). Article in Korean. Ministry of Food and Drug Safety. 2019 Dec 30. URL: https://nedrug.mfds.go.kr/bbs/58/3794/ [accessed 2022-09-07]
35. Famotidine. Drugs.com. URL: https://www.drugs.com/famotidine.html [accessed 2022-09-07]
36. Micromedex Products. IBM Corp. URL: https://www.micromedexsolutions.com/micromedex2/librarian/PFDefaultActionId/evidencexpert.DoIntegratedSearch?navitem=topHome&isToolPage=true# [accessed 2022-09-07]
37. Gupta N, Patel C, Panda M. Hepatitis following famotidine: a case report. Cases J 2009 Jan 27;2(1):89 [FREE Full text] [doi: 10.1186/1757-1626-2-89] [Medline: 19173722]
38. Wu Chou AI, Lu ML, Shen WW. Hepatotoxicity induced by clozapine: a case report and review of literature. Neuropsychiatr Dis Treat 2014 Aug 26;10:1585-1587 [FREE Full text] [doi: 10.2147/NDT.S67654] [Medline: 25210451]
39. Pise HN, Padwal SL. Diclofenac induced angioedema: a case report. Asian Journal of Pharmaceutical and Clinical Research 2015 Mar 01;8(2):4-5 [FREE Full text]
40. Jha AA, Bohra V, Behera V. Severe anaphylactic reaction to diclofenac. Med J Armed Forces India 2015 Jul;71(Suppl 1):S279-S281 [FREE Full text] [doi: 10.1016/j.mjafi.2014.08.015] [Medline: 26265858]
41. RS-ADR. URL: https://bioemr2.snubi.org:19108/ [accessed 2022-09-08]
42. Ye Y, Zhu YT, Xin XY, Zhang JC, Zhang HL, Li D. Efficacy of Chinese herbal medicine for tPA thrombolysis in experimental stroke: a systematic review and meta-analysis. Phytomedicine 2022 Jun;100:154072. [doi: 10.1016/j.phymed.2022.154072] [Medline: 35349833]
43. Seck B, Ndiaye MT, Diop A, Gaye C, Diouf A, Diagne FG, et al. The relevancy of patch testing in the exploration of the cutaneous side effects of herbal medicine. Our Dermatol Online 2021 Jan 2;12(1):19-23. [doi: 10.7241/ourd.20211.4]
44. Azizah N, Halimah E, Puspitasari IM, Hasanah AN. Simultaneous use of herbal medicines and antihypertensive drugs among hypertensive patients in the community: a review. J Multidiscip Healthc 2021 Feb 03;14:259-270 [FREE Full text] [doi: 10.2147/JMDH.S289156] [Medline: 33568913]

Abbreviations

ADR: adverse drug reaction
CVAD: controlled vocabulary–based ADR signal dictionary
EHR: electronic health record
FAERS: Food and Drug Administration Adverse Event Reporting System
FDA: Food and Drug Administration
ICD-10: International Classification of Diseases 10th Revision
LOINC: Logical Observation Identifiers Names and Codes
MedDRA: Medical Dictionary for Regulatory Activities
OMOP: Observational Medical Outcomes Partnership
PT: preferred term
PV: pharmacovigilance
RS-ADR: reference standard for adverse drug reaction
RWD: real-world data
RWE: real-world evidence
SIDER: Side Effect Resource
SNS: standard nursing statement
SNUH: Seoul National University Hospital
SOC: system organ class
WHO: World Health Organization
WHO-ART: World Health Organization Adverse Reactions Terminology

Edited by G Eysenbach; submitted 06.12.21; peer-reviewed by HW Han, C Jimeno, B Foroutan; comments to author 09.02.22; revised version received 29.04.22; accepted 14.07.22; published 06.10.22.

Please cite as:
Lee S, Lee JH, Kim GJ, Kim JY, Shin H, Ko I, Choe S, Kim JH
A Data-Driven Reference Standard for Adverse Drug Reaction (RS-ADR) Signal Assessment: Development and Validation
J Med Internet Res 2022;24(10):e35464
URL: https://www.jmir.org/2022/10/e35464
doi:10.2196/35464
PMID:36201386

Original Paper

Investigating Rewards and Deposit Contract Financial Incentives for Physical Activity Behavior Change Using a Smartphone App: Randomized Controlled Trial

David R de Buisonjé[1], MSc; Thomas Reijnders[1,2], PhD; Talia R Cohen Rodrigues[1], MSc; Santhanam Prabhakaran[3], MSc; Tobias Kowatsch[3,4,5], Prof Dr; Stefan A Lipman[6], PhD; Tammo H A Bijmolt[7], Prof Dr; Linda D Breeman[1], PhD; Veronica R Janssen[1,8], PhD; Roderik A Kraaijenhagen[9], MD, PhD; Hareld M C Kemps[10], MD, PhD; Andrea W M Evers[1,11], Prof Dr

[1]Health, Medical and Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden, Netherlands
[2]Department of Human-Centered Design, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands
[3]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[4]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[5]School of Medicine, University of St. Gallen, St. Gallen, Switzerland
[6]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands
[7]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands
[8]Department of Cardiology, Leiden University Medical Center, Leiden, Netherlands
[9]Hearts4People Foundation, Amsterdam, Netherlands
[10]Department of Cardiology, Máxima Medical Center, Veldhoven, Netherlands
[11]Medical Delta, Leiden University, Technical University Delft, Erasmus University, Delft, Netherlands

Corresponding Author:
David R de Buisonjé, MSc
Health, Medical and Neuropsychology Unit
Institute of Psychology
Leiden University
Wassenaarseweg 52
Room 2A22
Leiden, 2333 AK
Netherlands
Phone: 31 630665252
Email: d.r.de.buisonje@fsw.leidenuniv.nl

Abstract

Background: Financial incentive interventions for improving physical activity have proven to be effective but costly. Deposit contracts (in which participants pledge their own money) could be an affordable alternative. In addition, deposit contracts may have superior effects by exploiting the power of loss aversion. Previous research has often operationalized deposit contracts through loss framing a financial reward (without requiring a deposit) to mimic the feelings of loss involved in a deposit contract.

Objective: This study aimed to disentangle the effects of incurring actual losses (through self-funding a deposit contract) and loss framing. We investigated whether incentive conditions are more effective than a no-incentive control condition, whether deposit contracts have a lower uptake than financial rewards, whether deposit contracts are more effective than financial rewards, and whether loss frames are more effective than gain frames.

Methods: Healthy participants (N=126) with an average age of 22.7 (SD 2.84) years participated in a 20-day physical activity intervention. They downloaded a smartphone app that provided them with a personalized physical activity goal and either required a €10 (at the time of writing: €1=US $0.98) deposit up front (which could be lost) or provided €10 as a reward, contingent on performance. Daily feedback on incentive earnings was provided and framed as either a loss or gain. We used a 2 (incentive type: deposit or reward) × 2 (feedback frame: gain or loss) between-subjects factorial design with a no-incentive control condition. Our primary outcome was the number of days participants achieved their goals. The uptake of the intervention was a secondary outcome.

Results: Overall, financial incentive conditions (mean 13.10, SD 6.33 days goal achieved) had higher effectiveness than the control condition (mean 8.00, SD 5.65 days goal achieved; P=.002; ηp^2=0.147). Deposit contracts had lower uptake (29/47, 62%) than rewards (50/50, 100%; P<.001; Cramer V=0.492). Furthermore, 2-way analysis of covariance showed that deposit contracts (mean 14.88, SD 6.40 days goal achieved) were not significantly more effective than rewards (mean 12.13, SD 6.17 days goal achieved; P=.17). Unexpectedly, loss frames (mean 10.50, SD 6.22 days goal achieved) were significantly less effective than gain frames (mean 14.67, SD 5.95 days goal achieved; P=.007; ηp^2=0.155).

Conclusions: Financial incentives help increase physical activity, but deposit contracts were not more effective than rewards. Although self-funded deposit contracts can be offered at low cost, low uptake is an important obstacle to large-scale implementation. Unexpectedly, loss framing was less effective than gain framing. Therefore, we urge further research on their boundary conditions before using loss-framed incentives in practice. Because of limited statistical power regarding some research questions, the results of this study should be interpreted with caution, and future work should be done to confirm these findings.

Trial Registration: Open Science Framework Registries osf.io/34ygt; https://osf.io/34ygt

(J Med Internet Res 2022;24(10):e38339) doi:10.2196/38339

KEYWORDS

eHealth; behavior change; rewards; reward learning; financial incentives; deposit contracts; commitment contracts; physical activity; mobile phone

Introduction

Background

Since the beginning of time, humans have been developing tools and technologies that have made life easier. These technological advances have led to historically unprecedented levels of physical inactivity [1]. For example, currently, only 23% of adults in the United States meet the recommended guidelines for physical activity [2]. Although physical inactivity is linked to chronic disease and early death [3], increasing physical activity reduces the risk of chronic disease, has positive effects on mental health, and increases longevity [4]. Importantly, the positive effects of physical activity are observed not only for intense aerobic training but also for the mere number of steps taken in daily life [5,6]. Intervening on improving daily step counts has the advantage of being objectively measurable (compared with self-reports), low cost (compared with pharmaceutical treatment), and relatively easy to implement in daily life (compared with gym-based aerobic training), and as a result, it is also suitable for deprived, vulnerable, and older populations worldwide. Therefore, stimulating an increase in daily step counts appears to be a promising and feasible avenue to help humanity become healthier and happier and to live longer.

Although many people are aware of the benefits of physical activity and have positive intentions to be (more) physically active, achieving sufficient physical activity in daily life is not achieved by many [7]. The finding that positive intentions do not always translate into the desired behavior has been linked to the intention-behavior gap and has been found in a variety of (health) behaviors [8], including physical activity [7]. Insights from behavioral economics help explain the causes of the intention-behavior gap. A key finding is that people are present biased [9]. Present bias refers to the tendency of people to be more strongly driven by consequences in the here and now, rather than by the long-term consequences of their decisions. Consequently, people tend to procrastinate. Although differences among individuals exist, the general pattern found is one wherein "people grab immediate rewards and avoid immediate costs in a way our long-run selves do not appreciate" [10]. Present bias has been shown to apply to health behavior in general [11] and to physical activity specifically [12]. For example, people with a stronger present bias have lower levels of physical activity, arguably because they overweight the short-term and often negative consequences of physical activity (eg, increased heart rate and sweating) and assign a lower value (ie, discount) to the long-term positive consequences of physical activity (eg, longevity) [12]. Present bias, therefore, helps explain why despite having good intentions to achieve long-term health goals, people are prone to fall for immediate temptation.

Present bias also helps explain why introducing financial incentives might be suitable as an intervention strategy for health behavior change. Offering immediate financial incentives for healthy behavior takes advantage of the present bias by introducing a monetary benefit in the here and now. As such, people no longer have to *wait* for the delayed rewards of healthy behavior to emerge but instead are immediately rewarded. Indeed, meta-analyses and systematic reviews have shown that financial incentives are an effective tool for promoting (at least short-term) health behavior changes, such as improving diet [13], combating substance use [13], increasing physical activity [14,15], weight loss [13], smoking cessation [15,16], and increasing vaccination uptake [16]. Financial incentives are often added as a supplement to already active behavior change interventions, roughly doubling the odds of successful behavior change [15]. For physical activity, a recent meta-analysis (N=6074) on the effectiveness of financial incentives on step counts showed an average daily increase of approximately 600 steps (10%-15%) during active intervention [14].

Another relevant insight from behavioral economics is that people are loss averse [17]. This refers to the tendency of individuals to assign larger weight to potential losses associated with their behavior than to potential gains. Losses and gains are defined with respect to a reference point; for example, individuals' current status quo, their expectations, or goals [17]. Loss aversion and reference points have been shown to be

important in health-related decision-making [18] and might lead to suboptimal decision-making for physical activity if it causes people to outweigh what they might lose by being physically active (eg, time and energy) over what they might gain (eg, satisfaction after a workout). Furthermore, loss aversion is often used to motivate financial incentive designs that involve potential losses rather than rewards only [19,20], such as deposit contracts.

Deposit contracts are a specific form of financial incentive wherein people deposit their own money and can earn it back contingent on behavior change [21]. There are several real-world commercial products (eg, Waybetter [22] and Stickk [23]) with deposit contracts that have proven to be commercially viable and claim to help people change their behavior. While rewards involve the introduction of a pleasant stimulus to increase behavior (ie, positive reinforcement), deposit contracts involve the alleviation of an aversive stimulus (avoiding loss of money) to increase behavior (ie, negative reinforcement) [24]. Deposit contracts offer several advantages over reward-based incentives. First, although both rewards and deposit contracts bring an incentive into the present, a deposit contract brings a risk of loss into the present and thus should be more effective because it capitalizes on loss aversion [19]. Second, the use of reward-based financial incentives for physical activity imposes a significant cost (eg, approximately US $1.50 per day per person, see the study by Mitchell et al [14]), whereas the use of deposit contracts introduces (partial) cost sharing by recipients. Such cost sharing may be desirable, for example, to employers promoting physical activity among employees [25]. Moreover, while rewarding people for behavior that others perform without receiving rewards might be considered unfair, having people voluntarily deposit their own money avoids this ethical concern [26].

Existing evidence indicates that deposit contracts are effective in helping people lose weight [26], stop smoking [19,27], and increase physical activity [20,21,24,28-30]. However, the voluntary uptake of deposit contracts is generally low [19,31]. In fact, some authors suggest that those who would benefit the most from interventions using incentives with potential losses are not likely to enter into them [32,33]. However, comparing the evidence on the uptake and effectiveness of deposit contracts for physical activity among studies is complicated, as operationalizations differ substantially. In particular, 3 different types of deposit contracts can be distinguished. First, in line with their potential to promote cost sharing, several authors have used completely self-funded deposit contracts [31,34]. Without the potential for financial gain, such self-funded deposit contracts involve only losses compared with the status quo. Second, uptake of deposit contracts is often encouraged through "matching" individuals' contribution into the deposit scheme or combining deposits with a reward-based incentive [19,35,36]. Such matched deposit contracts thus involve both potential gains and losses compared with the status quo. Third, some authors have used loss framing to mimic the feelings of loss involved in a deposit contract without actually requiring individuals to put their own money at risk [20,24]. For example, in a loss-framed condition, Patel et al [20] promised respondents US $42 up front of which they could then lose US $1.40 for every day they did not attain physical activity goals. This loss-framed condition proved more effective in promoting physical activity compared with a gain-framed condition in which respondents simply earned US $1.40 for every day they attained physical activity goals. However, participants in all conditions of this study faced no actual losses, but in fact were making gains compared with their preintervention status quo.

This Study

In this study, we investigate the impact of deposit contracts on increasing physical activity by disentangling the effects of incurring actual losses (through self-funding) and loss framing. We will use an actual deposit contract (ie, a stick) that requires participants to make a deposit of their own money before the intervention starts and compare this with receiving a reward (ie, a carrot) of equivalent size. In line with the study by Adams et al [37], we refer to this as the *direction* of incentives. Furthermore, we will investigate whether loss framing (compared with gain framing) enhances the effectiveness of both reward and deposit contract incentives. First, we expect that, overall, incentive conditions are more effective than an active no-incentive control condition (H1). Second, we hypothesize that deposit contracts will have lower uptake than regular rewards (H2); however, deposit contracts are expected to be more effective than regular rewards for those that partake in the intervention (H3). In addition, we hypothesize that loss framing an incentive will increase effectiveness compared with gain framing (H4). Finally, we propose that incentives in which both direction of the incentive and framing of the incentive are loss congruent (ie, loss-framed deposit contracts) are most likely to invoke loss aversion and are therefore especially effective in promoting physical activity (H5).

Methods

Participants

We recruited healthy participants aged between 18 and 30 years through a university research participation system (SONA), flyers on campus, and posts on social media. Participants had to be willing to improve their physical activity, own a smartphone, and be proficient in English. A priori sample size calculations with G*Power [38] suggested a minimum sample size of 199 for detecting a between-conditions difference in effectiveness with a medium effect size (f=0.20), 80% power, and an α of .05 (analysis of covariance [ANCOVA] with 5 groups). On the basis of a similar research [39] that showed a relatively high dropout rate between recruitment and participation, we assumed a dropout rate of 20% and aimed to recruit 240 eligible participants. Participants were excluded if they reported any medical condition that could hinder their physical activity (based on their response to the Physical Activity Readiness Questionnaire) [40]. A detailed description of the flow of participants through the study, including reasons for exclusion and dropout, is provided in Multimedia Appendix 1. All the participants who completed the study had a chance to win 1 of 3 grand prizes (3 Fitbit devices worth €100 [at the time of writing: €1=US $0.98]) and 1 of 50 small prizes (50 webshop vouchers worth €10) in a raffle. Participants who were

first-year psychology students at Leiden University additionally received research credits (needed to complete their first year).

Ethics Approval

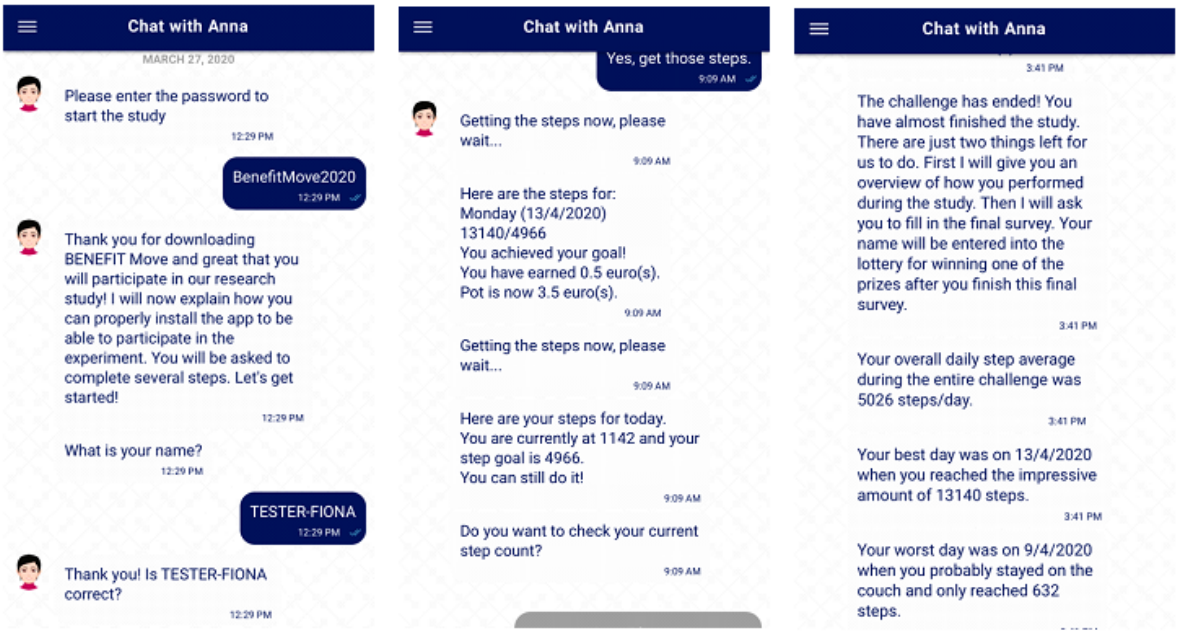
We obtained informed consent before the start of the study. This study was approved by the Psychology Research Ethics Committee of Leiden University (2020-02-24-T. Reijnders-V2-2089), and the study protocol was preregistered on the Open Science Framework [41].

Materials

The intervention for this study was delivered entirely on the web via the Benefit Move app, which the participants downloaded on their smartphones. The Benefit Move app was implemented using MobileCoach [42,43], an open-source software platform for smartphone-based and chatbot-delivered behavioral interventions (eg, study by Kowatsch et al [44]) and ecological momentary assessments (eg, study by Tinschert et al [45]). MobileCoach was developed by the Centre for Digital Health Interventions at Eidgenössische Technische Hochschule Zürich and the University of St. Gallen in Switzerland [46]. The Benefit Move app had two main functions: (1) objectively measuring physical activity and (2) communicating with the participant.

To measure physical activity, the Benefit Move app asked the participants for permission to retrieve step counts from existing health apps already installed on their smartphones. Most smartphones have a gyroscope-based pedometer or location-tracking device integrated to record movements made while the phone is being carried. Algorithms recode the raw data from these sensors into an estimated step count, which is then stored in the database of apps, such as Apple Health and Google Fit. Depending on the operating system, Benefit Move would pull data from either Google Fit [47] for Android or Apple's Health Kit [48] for iOS. Overall, out of 126 participants, 67 (53.2%) used Apple iOS devices and 59 (46.8%) used Android devices. The percentage of Apple iOS users ranged from 41.1% to 69.6% across conditions and was considered to be spread evenly across conditions. Both of these apps showed good validity for measuring step counts [49,50]. The Benefit Move app retrieved these data to provide a tailored step goal at the start of the intervention and to record step counts during the intervention. During the intervention phase, at any given time, the participant could click a button to retrieve the up-to-date step count at that moment. In addition, to communicate with the participant, an automated digital coach (chatbot) sent daily prompts to provide the participant with feedback about goal progress, their accumulated financial earnings or losses, and a trigger to click the button for step count retrieval (Figure 1 provides an impression of the app).

Figure 1. Impression of the Benefit Move app.



Measures

Baseline Survey

The baseline survey was administered during onboarding in the app to obtain basic demographic information such as sex, birth year, nationality, country of residency, education level, employment status, subjective estimation of income relative to peers, and subjective estimation of weight status (Multimedia Appendix 2 provides an overview of the survey items).

Final Survey

The final survey was administered after the intervention was completed. First, as a sensitivity check, we asked the participants whether they carried their smartphone with them more often because of the intervention (Multimedia Appendix 3 provides an overview of the final survey items). Furthermore, we asked the participants if they cheated the intervention but assured them that their answer would not impact the payout of incentives. We also performed a contamination check to explore whether participants were aware of the condition that others were

assigned. Because the intervention coincided with the worldwide COVID-19 pandemic, we included several items to assess its impact on our study. First, we assessed whether participants experienced influenza-like symptoms, whether these symptoms led them to be less physically active, and, in general, whether they engaged in less physical activity owing to the COVID-19 pandemic. Furthermore, we administered the Generalized Anxiety Disorder-7 [51], a brief 7-item measure that assesses generalized anxiety symptoms that could be related to the COVID-19 pandemic. Finally, as a manipulation check, we included 2 items (answered on a 10-point Likert scale from 1=totally disagree to 10=totally agree) that asked whether participants experienced a feeling of loss during the intervention ("I felt that I was losing money if I did not increase my step count") and whether they experienced goal commitment ("I felt strongly committed to the goal of increasing my step count").

Procedure

After recruitment, all the participants were put on a waitlist before they received the screening survey and informed consent. One week before the start of the intervention, participants completed the screening survey with the inclusion and exclusion criteria and provided digital informed consent. Thereafter, eligible participants received a URL to the iOS or Android app stores where they could download the Benefit Move intervention app and install it on their smartphone. Once the participants installed the app, they were asked to complete onboarding in the app within 2 days. Thereafter, participants were sent a link to the survey platform LimeSurvey that opened within the Benefit Move app. Here, they filled in the baseline survey (for more details, see the Baseline Survey section) and then returned to the app after completion. Participants were excluded from the study if they did not complete the onboarding process and baseline survey before the start of the intervention.

After participants completed the baseline survey, they received a tailored step goal based on their 7-day historic daily step average that was retrieved through Google Fit or Apple Health. Retrieving step counts for 7 consecutive days should accurately estimate habitual activity levels of individuals [52], and providing an individualized and realistic goal should increase intervention effectiveness [14]. A limitation to using a 7-day historic step count is that meteorological factors could impact baseline levels of activity [53]. If historic data were available, the participant was assigned a goal that was 120% of the historic daily step average. For example, someone who, in the 7 days before goal setting, took an average of 5000 steps per day would automatically receive a 6000 steps daily step goal. If no historic data were available, the participant was assigned a default step goal of 10,000 steps per day because it is an often-used guideline for sufficient physical activity [54].

All the participants started simultaneously with the 20-day intervention on Monday, March 30, 2020, at 9 AM. Owing to the COVID-19 pandemic, a partial lockdown was issued by the Dutch government on March 15, 2020. Onboarding for this study (and retrieval of 7 days of historic step counts) was performed from March 23, 2020, until the active study phase started on March 30, 2020. Therefore, it is possible that the estimates of the baseline activity were lower than normal. Each day during the 20-day intervention, the participants received a push notification at 9 AM. This notification prompted them to click a button to retrieve their step count performance of the previous day and get an update on the progress for the current day. If the user skipped doing this for several days but then responded and requested an update, the feedback for multiple days was given in separate consecutive messages, with a separate update message per day. The feedback per day consisted of the achieved step count compared with the daily step goal, a conclusion about whether the goal was achieved or not, the money that was earned or lost on that day, and the running total of earnings or losses during the entire intervention (Figure 1 provides an example). On the basis of their study conditions, participants received different instructions at the start of the intervention and received different feedback messages during the intervention.

Study Conditions

We used a 2 incentive direction (reward or deposit) × 2 feedback frame (gain or loss) design with an additional control condition. The participants were automatically randomized to these 5 conditions by the app.

Condition 1: Control Condition

Participants received an active basic intervention with a tailored goal and daily feedback on their goal progress without a financial incentive or specific framing of feedback.

Condition 2: Reward and Gain Frame Condition

After having been assigned their step goal, participants were informed that they would receive a monetary reward of a maximum of €10 for achieving their step goals during the intervention (the incentive amount of €10 was determined in a pilot study during which we sent a short survey to 26 students to assess what incentive amount they would find stimulating and acceptable). More specifically, to create a gain frame, they were informed that there was an empty pot at the start of the intervention and that for every successful goal achievement, they would receive €0.50 that would be added to the pot. If they were not successful, nothing would be added to the pot. After their condition was explained to the participants, we explicitly asked them if they wanted to participate in this challenge (this is especially relevant for participants in the deposit conditions, as they will be asked to make a monetary payment to the experiment). After they explicitly agreed to the specific challenge that was presented to them, the participants were instructed to wait until the intervention started the next Monday morning.

Condition 3: Reward and Loss Frame Condition

After having been assigned their step goal, participants were also informed that they would receive a monetary reward of a maximum of €10 for achieving their step goals during the intervention. However, to create a loss frame, and in contrast to the gain frame condition, they were informed that there was a full pot with €10 at the start of the intervention and that for every goal failure €0.50 would be deducted from the pot. If they were successful, nothing would be deducted from the pot.

Condition 4: Deposit and Gain Frame Condition

After having been assigned their step goal, participants in the deposit and gain frame condition were asked to deposit €10 of their own money via bank transfer to improve their commitment to the challenge. In all cases, the full amount was refunded after the intervention, but participants were unaware of this and were informed that the amount they would get back would depend on their performance during the intervention. More specifically, they were informed that there was an empty pot at the start of the intervention and that for every successful goal achievement, €0.50 would be added to the pot. If they were not successful, nothing would be added to the pot. The final amount of the pot would be the amount of their deposit that would be returned to them after the intervention.

After their condition was explained to them, we explicitly asked the participants if they wanted to participate in this challenge. When participants agreed to participate, they were sent a digital payment request via "Tikkie" (a direct digital payment URL) in the app. By clicking on this payment request, they directly transferred €10 of their own funds to the experiment bank account. Participants who could not use this automated system were able to transfer the required amount manually to the experiment bank account. The experiment bank account was monitored closely, and when a deposit payment was received, we confirmed this to the participant through the intervention app. If no payment was received, participants were automatically reminded via push messages, SMS text messages, telephone calls, and email reminders. Participants were excluded if deposit payments were not confirmed 12 hours before the start of the intervention. After confirming the received deposit payment, we instructed the participants to wait until the intervention started the next Monday morning.

Condition 5: Deposit and Loss Frame Condition

Participants in this condition followed the same overall procedure as the participants in the deposit and gain frame condition did. However, to create the loss-framed feedback, they were informed that there was a full pot of €10 at the start of the intervention and that for every goal failure, €0.50 would be deducted from the pot. If they were successful, nothing would be deducted from the pot. The final pot amount was the amount of their deposit that we promised to return after the intervention.

Debriefing

After the participants completed the 20-day intervention, they received a summary of their performance in the challenge. In the 4 experimental conditions, the participants were additionally informed about their incentive earnings and told that they would receive this money (back) into their bank account as soon as possible. Thereafter, the participants were sent a link to the survey platform LimeSurvey that opened within the Benefit Move app. Here, they filled in the final survey (for more details, see the Final Survey section) and returned to the intervention app after completion. Participants were then debriefed about their condition; the other conditions and the deceptive element around their deposit were revealed. All payments to the participants were made within 2 weeks after the experiment ended.

Statistical Analysis

The primary outcome (continuous) was the effectiveness. This was measured through the mobile registration of step count data and defined as the number of days (0-20) the goal was achieved. The secondary outcome (binary) was the uptake of the intervention and defined as explicitly agreeing to participate in the challenge and paying the deposit (if required).

We report results on the effectiveness based on a restricted sample that only included participants who retrieved steps on at least one intervention day and who received a tailored step goal. We excluded participants who received a default goal, because in hindsight, these participants were confronted with a goal that was unachievable (Multimedia Appendix 4 provides an overview of analyses where these participants were included). Furthermore, we report the main analyses for effectiveness based on models that include baseline step counts as a covariate. The pattern of the results was similar, but the models gained accuracy by including the covariate. Data analysis was performed using SPSS statistics for Mac (version 28; IBM Corp). We dealt with missing cases by using pairwise exclusion and used the standard $P<.05$ criterion for determining statistical significance. For ANOVA and ANCOVA, we considered an effect size small when $\eta p^2>0.01$, medium when >0.06, and large when >0.14 (Cohen [55]). For chi-square, we considered an effect size small when Cramer $V>0.1$, medium when >0.3, and large when >0.5.

Hypothesis Testing

Hypothesis 1: Effectiveness of Incentive Conditions Compared With the Control Condition

First, we performed an ANCOVA with baseline steps as a covariate in which we compiled incentive conditions to compare all incentive conditions combined (mean of conditions 2-5) to the control condition (ie, condition 1). Second, we performed an ANCOVA with baseline steps as a covariate and effectiveness as the dependent variable to separately compare incentive conditions (ie, conditions 2-5) to the no-incentive control condition (ie, condition 1). The ANCOVA was performed with factor "condition" with 5 levels (conditions 1-5). We compared each incentive group separately to the control condition with four planned contrasts: 1=control versus deposit and gain, 2=control versus deposit and loss, 3=control versus reward and gain, and 4=control versus reward and loss.

Hypothesis 2: Uptake of the Intervention

We performed a chi-square test of independence to investigate whether the uptake was lower for deposit contracts (ie, conditions 4 and 5) compared with regular rewards (ie, conditions 2 and 3).

Hypothesis 3 to 5: The Effect of Incentive Direction and Feedback Framing on Effectiveness

We performed a 2-way ANCOVA with baseline steps as a covariate. Effectiveness was the dependent variable, and the model contained 2 factors: incentive direction (deposit or reward) and feedback frame (loss or gain). In the model, we specified both the main effects of the factors (H2 and H3) and their interactions (H4).

Results

Descriptives

In total, we analyzed the data on the uptake of participants (N=126) with a mean age of 22.7 (SD 2.84) years of which 68.2% (86/126) identified as female. Most participants had the Dutch nationality (69/126, 54.8%), approximately half (60/126, 47.6%) were students, most reported to have an income similar to their peers (71/126, 56.3%), and most considered themselves to have an appropriate body weight (89/126, 70.6%). After their condition was explained to them, 11 participants explicitly refused the challenge, 7 participants did not pay their deposit in time, and 12 participants did not retrieve steps on any day of the intervention. Therefore, the data from 96 participants were available for the analysis of effectiveness, and the data from 65 participants remained after exclusion of nontailored goals (see the Methods section for rationale). Table 1 provides more details on the characteristics of the full sample that was analyzed for uptake and the subsample that was analyzed for effectiveness.

Table 1. Sample characteristics of the full sample and the subsample that was analyzed for effectiveness.

Variable	Full sample (N=126)	Subsample effectiveness (N=65)
Age (years), mean (SD)	22.7 (2.84)	22.2 (2.53)
Sex, n (%)		
Male	40 (31.7)	13 (20)
Female	86 (68.3)	52 (80)
Nationality, n (%)		
Dutch	69 (54.8)	40 (61.5)
German	20 (15.9)	10 (15.4)
Other	37 (29.4)	15 (23.1)
Work , n (%)		
Student without a job	54 (42.8)	33 (50.8)
Student with a job	6 (4.8)	1 (1.5)
Working part time	14 (11.1)	6 (9.2)
Working full time	45 (35.7)	21 (32.3)
Do not want to answer	7 (5.6)	4 (6.2)
Income, n (%)		
Less than my peers	15 (11.9)	9 (13.8)
Same as my peers	71 (56.3)	39 (60)
More than my peers	20 (15.9)	9 (13.8)
Do not want to answer	20 (15.9)	8 (12.3)
Weight (kg) , n (%)		
Underweight	3 (2.4)	1 (1.5)
A bit underweight	7 (5.6)	4 (6.2)
Appropriate weight	89 (70.6)	48 (73.8)
A bit overweight	19 (15.1)	9 (13.8)
Overweight	7 (5.6)	2 (3.1)
Do not want to answer	1 (0.8)	1 (1.5)

Hypothesis Testing

Hypothesis 1: Effectiveness of Incentive Conditions Compared With Control Condition

First, a 1-way ANCOVA with baseline steps as a covariate showed that, overall, incentive conditions (mean 13.10, SD 6.33 days goal achieved) had higher effectiveness than the control condition (mean 8.00, SD 5.65 days goal achieved; $F_{1,62}$=10.72; P=.002; ηp^2=0.147). Furthermore, to test specific contrasts, a second 1-way ANCOVA with baseline steps as a covariate showed that the factor condition was related to the effectiveness of the intervention ($F_{4,59}$=5.48; P<.001; ηp^2=0.271). Participants in the control condition achieved their step goal on a mean of 8.00 (SD 5.65) days. Planned contrasts indicated that this was significantly less than that in the participants in reward and gain condition (mean 13.30, SD 5.49 days goal achieved; P=.003; SE 1.86). Furthermore, this was also significantly less than that of participants in the deposit and gain condition (mean 17.40, SD 6.17; P<.001; SE 2.25). We did not find a significant

difference between the control condition and the reward and loss condition (mean 10.00, SD 7.01 days goal achieved; *P*=.23; SE 2.19). No significant difference was found between the control condition and the deposit and loss condition (mean 11.29, SD 5.16 days goal achieved; *P*=.19; SE 2.53). Owing to indications that normality of the dependent variable was violated, we performed a Kruskal-Wallis test to check the robustness of these findings. We only found a significant contrast between the control condition and the deposit and gain condition (*P*=.001, adjusted with Bonferroni correction). There was no evidence of a significant difference for the other contrasts.

Hypothesis 2: Uptake of the Intervention

Uptake of the intervention was defined as explicitly agreeing to participate in the challenge and paying the deposit (if required). A chi-square test of independence showed that requiring a deposit decreased the uptake of the intervention (N=97; χ^2_1=23.5; *P*<.001; Cramer *V*=0.492). In the reward conditions, 100% (50/50) of the participants accepted the intervention compared with 62% (29/47) in the deposit conditions (Table 2 provides a descriptive overview of the results). We explored whether those with uptake differed from those with no uptake but were underpowered for these analyses and accordingly found no differences in demographic data (sex, income, weight status, and age) or other baseline characteristics (goal type, self-efficacy, risk proneness, self-control, autonomous motivation, extrinsic motivation, and historic step count).

Table 2. Descriptive overview of the results.

Variable	Condition					Total (N=126)
	Control (n=29)	Reward and gain frame (n=32)	Reward and loss frame (n=18)	Deposit and gain frame (n=23)	Deposit and loss frame (n=24)	
Uptake, n (%)	29 (100)	32 (100)	18 (100)	15 (65)	14 (58)	108 (86)
Explicit refusal, n (%)	0 (0)	0 (0)	0 (0)	4 (17)	7 (29)	11 (9)
Deposit not paid, n (%)	N/A[a]	N/A	N/A	4 (17)	3 (12)	7 (6)
Steps never retrieved, n (%)	2 (7)	4 (12)	3 (17)	0 (0)	3 (12)	12 (10)
Goal type, n (%)						
Tailored goals	18 (62)	21 (66)	11 (61)	17 (74)	14 (58)	81 (68)
Default goals 10,000	11 (38)	11 (34)	7 (39)	6 (26)	10 (42)	45 (36)
Assigned step goal, mean (SD)	6189 (3604)	6384 (3700)	6992 (3111)	5960 (3544)	7714 (3724)	6602 (3574)

[a]N/A: not applicable.

Hypothesis 3 to 5: Effect of Incentive Direction and Feedback Framing on Effectiveness

A 2-way ANCOVA with baseline steps as a covariate showed no main effect of incentive direction ($F_{1,43}$=1.98; *P*=.17; ηp^2=0.044), indicating that deposits (mean 14.88, SD 6.40 days goal achieved) were not more effective than rewards (mean 12.13, SD 6.17 days goal achieved). We did find a main effect of feedback framing ($F_{1,43}$=7.91; *P*=.007; ηp^2=0.155), indicating that loss frames (mean 10.50, SD 6.22 days goal achieved) were significantly less effective than gain frames (mean 14.67, SD 5.95 days goal achieved). Finally, the interaction effect of incentive direction×feedback framing was not significant ($F_{1,43}$=1.16; *P*=.29; ηp^2=0.026), indicating that feedback framing did not have a different effect on deposit conditions compared with reward conditions. Table 3 provides a descriptive overview of the results for each arm of the experiment.

Furthermore, to test the robustness of these findings, we additionally performed a Kruskal-Wallis test. For the main effects, we performed 2 separate tests, one for each factor from the 2-way ANOVA. However, the interaction effect could not be tested with this alternative method. Consistent with the results of the 2-way ANCOVA, we found that incentive direction was not significantly related to effectiveness (*P*=.06), but feedback framing was significantly related to effectiveness (*P*=.03). Additional checks to test the sensitivity of the main findings are reported in Multimedia Appendix 5.

Table 3. Descriptive overview of results for participants with tailored goals.

Variable	Condition, mean (SD)					Total (N=65), mean (SD)
	Control (n=17)	Reward and gain frame (n=20)	Reward and loss frame (n=11)	Deposit and gain frame (n=10)	Deposit and loss frame (n=7)	
Baseline step count	3406 (1982)	3868 (2673)	4232 (2056)	4036 (3187)	3472 (1537)	3792 (2347)
Assigned step goal	4087 (2378)	4642 (3207)	5078 (2467)	4843 (3825)	4166 (1844)	4550 (2816)
Intervention step count	3130 (2466)	5071 (2783)	4763 (2105)	6395 (4526)	3993 (2464)	4599 (3025)
Days goal achieved	8.00 (5.65)	13.30 (5.49)	10.00 (7.01)	17.40 (6.17)	11.29 (5.16)	11.77 (6.52)

Effect of the Manipulations on Experienced Feelings of Loss and Goal Commitment

To check the effect of our manipulations, we analyzed the effects of incentive direction and feedback framing on feelings of loss and goal commitment. We performed 2 separate 2-way ANOVAs (one for feeling of loss and one for goal commitment) with factor incentive direction (deposit or reward) and factor feedback frame (loss or gain). The model included both main effects and their interactions. The first ANOVA, with feeling of loss as the dependent variable, showed a significant effect of incentive direction ($F_{1,41}$=19.66; P<.001; ηp^2=0.324). Deposit contracts (mean 7.19, SD 2.23) resulted in stronger feelings of loss compared with rewards (mean 4.21, SD 2.19). However, feedback framing did not influence the feeling of loss, and we did not find a significant interaction. The second ANOVA, with goal commitment as the dependent variable, showed a significant effect of feedback framing ($F_{1,41}$=4.95; P=.03; ηp^2=0.108). Loss-framed incentives (mean 5.24, SD 3.11) resulted in weaker goal commitment compared with gain-framed incentives (mean 7.14, SD 2.37). However, incentive direction did not influence goal commitment, and we did not find any interaction.

Discussion

Principal Findings

This study found that financial incentives increase intervention effects compared with an active no-incentive control condition. Furthermore, as expected, the results showed that self-funded deposit contracts for physical activity have a lower uptake than regular reward incentives. However, in contrast to our hypothesis, we did not find deposit contracts to be more effective than reward incentives, but they were also not less effective and have important benefits for large-scale implementation. An important unexpected finding was that loss framing decreased the effectiveness of the intervention compared with gain framing. This finding is in contrast to the existing literature and seems to provide the first preliminary evidence that for improving physical activity with financial incentives in a healthy population, loss framing is less effective than gain framing.

First, the finding that financial incentive conditions were more effective than an active no-incentive control condition is in line with the results from meta-analyses [14-16]. Compared with participants in the control condition, participants who received a financial incentive were shown to reach about 5 more daily step goals (and took about 2000 steps more per day) during the 20-day intervention. This is a large and clinically relevant effect with a mortality-reducing potential [5,6]. We explain this finding through the idea that financial incentives capitalize on the present bias and introduce an immediate monetary incentive for being physically active.

Second, we found that the uptake of deposit contracts was lower than that of regular rewards. This finding is in line with the work by Halpern et al [19] on deposit contracts for smoking cessation. A common sense explanation for this finding is that people are more open to an intervention where they stand to gain something (ie, a reward) than where they stand to lose something (ie, their own money). The same aversion to losses that is thought to increase effectiveness might deter people from entering into a deposit contract. In fact, this tension between effectiveness and uptake has been recognized before [56]. Furthermore, although we simplified all steps in the payment process, it could be that the logistical barrier of having to provide a monetary deposit deterred some individuals, regardless of whether they dismissed the concept of deposit contracts per se. Finally, it is important to understand which people are most likely to accept and reject a deposit contract intervention. For example, it has previously been suggested that individuals who recognize their challenges while resisting temptation (ie, sophisticates) might be open to using deposit contracts [56]. Future research should use a self-funded deposit contract and investigate the moderators of uptake to shed light on which subgroups are best reached.

Third, in contrast to our hypothesis, deposit contracts were not more effective than regular reward incentives. We expected, in line with others, that deposit contracts would invoke loss aversion and therefore would be more effective than regular rewards. Our analyses indeed showed that deposit contracts resulted in stronger feelings of loss than rewards did, but this did not result in higher effectiveness. Our results are in contrast to those reported for smoking cessation by Halpern et al [19]. Possibly, for physical activity, deposit contracts are not more effective than rewards. Another explanation might be that participants perceived the stakes in our study as low and therefore were not averse to potentially losing their deposits. This would be in line with the work by Mukherjee et al [57] who found that for high stakes, participants rated losses more impactful than gains (ie, loss aversion), but for low stakes, this tendency reversed, and gains were rated as more impactful than losses. It is possible that subjective judgments by our participants rated the incentive as low stakes and therefore deposit contracts were not more effective than rewards. Future

work should investigate deposit contracts and rewards of varying sizes to determine the potential tipping points at which deposit contracts are superior to rewards and when this is reversed. In addition, it is possible that deposit contracts are superior to rewards (the descriptive means were in the expected direction), but we did not have enough statistical power to detect a significant difference. More fully powered studies that investigate self-funded deposit contracts for physical activity are needed to draw firmer conclusions on this point. Existing studies in the domain of physical activity either operationalized deposit contracts differently using loss framing [20,24] or were also not powered [21,28-30] to provide a clear answer to this question.

Finally, unexpectedly, we found that loss framing decreased the effectiveness of the intervention compared with gain framing. In line with the study by Patel et al [20], we expected that framing an incentive as a loss would activate loss aversion and therefore increase effectiveness compared with gain framing an incentive. However, our analyses showed that loss framing did not increase feelings of loss compared with gain framing. Thus, it appears that our attempt at shifting participants' reference point was unsuccessful. We did find that loss framing decreased feelings of goal commitment, which might explain why the effectiveness of loss frames was lower than that of gain frames. Our results contradict the findings of Patel et al [20] who showed that loss-framed incentives were more effective than gain-framed incentives. However, Patel et al [20] studied university employees who are obese, with a BMI >27, whereas our sample consisted of healthy university students. Possibly, a difference in regulatory fit related to differences in the study sample might explain this discrepancy. Regulatory fit is when the persuasiveness of a health message is increased when its frame is congruent with the regulatory orientation of the individual [58]. Regulatory focus theory discerns 2 modes of regulatory orientation: promotion focus and prevention focus. Although people with a promotion focus aim for desired end states, people with a prevention focus aim for avoiding undesired end states [58]. Perhaps, adults who are obese are more focused on avoiding obesity-related health problems, and therefore have a stronger prevention focus when increasing physical activity. This could lead them to respond better to a loss-framed incentive (in which losing money is prevented) because of a greater experienced regulatory fit. By contrast, perhaps healthy students have a stronger promotion focus (on becoming more fit rather than avoiding health problems) and therefore respond better to a gain-framed incentive. Whether the regulatory fit effect also applies to incentive framing (and not only to framing of persuasive health messages) is an interesting avenue for future research. Future research should measure regulatory orientation and investigate the possible interactions with different incentive frames.

Strengths and Limitations

An important strength of this study is that we used a self-funded deposit contract that required participants to make a monetary deposit before the intervention started. This allowed us to compare the effects of self-funded deposit contracts with those of loss frames. Another strength is that we used objective registrations of step counts and did not rely on self-reported estimations of physical activity. Finally, the app automatically provided participants with tailored goals based on their historical step counts, thus creating a personalized intervention experience. However, requiring a deposit beforehand also resulted in a lower uptake of the deposit contract conditions. As a result, the deposit requirement may have filtered out people who lacked motivation, thus leading to an overestimation of effectiveness in the deposit contract conditions. Consequently, caution is warranted when interpreting the effectiveness of the deposit contract conditions. Another limitation of our study is that high dropout before onboarding, unbalanced allocation, lack of uptake in the deposit contract conditions, and the exclusion of nontailored goals decreased the statistical power of our analyses. Limited statistical power might have especially affected the findings for specific analyses on effectiveness such as when we compare deposit contracts with regular rewards or loss frames with gain frames. Therefore, the results of this study should be interpreted with caution, and future work should be done to confirm these findings. Furthermore, before onboarding, participants read the informed consent form, which mentioned that the study possibly required them to deposit €10 of their own money. Mentioning this possibility was important for informed consent but may have deterred some participants from participating before they onboarded in the app. It is possible that this biased our analysis of uptake and that the actual uptake of deposit contracts is lower than our analyses suggest. In addition, although we propose that objective measures of physical activity are superior to subjective self-reports, an important criticism of pedometer-based intervention research is that it is impossible to differentiate an increase in step count from an increase in pedometer wear time [59]. In our case, participants in the gain-framed conditions reported having carried their smartphone more often than they normally do (Multimedia Appendix 5), and this might partly explain why gain-framed conditions were more effective than loss-framed conditions. Furthermore, a relatively high proportion of the participants (45/126, 35.7%) did not have historical step data available on their smartphones. These people were assigned a default goal (10,000 steps per day) that was unachievable in hindsight. Although 10,000 steps per day is often used as a goal in commercial physical activity trackers and apps, this already exceeds the guidelines for sufficient physical activity, which translates to approximately 7000 to 8000 steps per day [60]. Future research with a similar goal-setting module should assign more achievable default goals when the goals cannot be tailored. In our sample, the mean baseline step count of participants with historical data was approximately 3800 steps per day. On the basis of a meta-analysis of financial incentive intervention effects, we suggest that step goals should not exceed baseline levels by >20% to 30% [14]. In addition, the intervention was launched in March 2020, and during this period, the first COVID-19 lockdown measures in the Netherlands were implemented. Although this probably impacted all conditions equally, a large part (51/65, 78%) of the sample reported having been less physically active than they normally were because of the situation around COVID-19. As a result, it is possible that the estimates of baseline activity were lower than normal; therefore, the intervention led to stronger improvements than would be found under normal circumstances. Furthermore, our

sample consisted of predominantly healthy, young, female students at universities. Although we purposefully recruited a homogenous sample to increase internal validity, the external validity of our findings is therefore restricted. Older or more chronically ill populations might respond differently to this type of intervention. Finally, we only investigated short-term effects during a 20-day intervention period. Therefore, we are unable to answer questions about the long-term effectiveness of the different incentive directions and incentive frames that we tested. Future work with longer intervention durations should be done to study how rates of goal achievement (and step counts) vary over time during and after the intervention.

Implications

An important theoretical contribution of this study is that we did not replicate the finding that loss-framed financial incentives are more effective than gain-framed financial incentives for increasing physical activity [20]. By contrast, our results show that gain-framed incentives are more effective. Although we are unable to ascertain what has produced this effect, by itself it provides evidence that (perceptions of) losses are not always more impactful than (perceptions of) gains. Rather, it supports the argument made by Gal and Rucker [61] that loss aversion is a context-dependent tendency with boundary conditions, instead of a ubiquitous phenomenon. This finding also has implications for those who want to implement loss-framed financial incentives in practice. Because our results show that loss frames might hurt incentive effectiveness, we warn against implementing them in practice without further research on their boundary conditions. Finally, we were unable to show that deposit contracts were more effective than rewards, but they were also not less effective. Considering that deposit contracts are (partially) self-funded makes them attractive for large-scale implementation. However, before deposit contracts can be implemented on a large scale, it is important to further understand which subgroups are not reached by them. Although to the best of our knowledge the relationship between income and uptake of deposit contracts has not yet been studied, one can imagine that people with lower incomes might reject a deposit contract because they are less able to deposit a sum of their own money. This could cause vulnerable key subgroups (eg, people with lower socioeconomic status or cardiovascular disease) not to be reached by a deposit contract intervention. Possibly, this issue could be overcome by offering income-dependent deposit sizes or allowing participants to freely choose an amount that is motivating but that does not cause financial harm when lost [26].

Conclusions

Although this study was underpowered and the results have to be interpreted with caution, we have shown that deposit contracts have lower uptake than rewards but appear to have (at least) comparable effects on physical activity. Loss framing an incentive might undermine effectiveness, and we therefore urge for more research before implementing them in practice. Deposit contracts might be a promising tool for behavior change; however, more research is needed on who is willing to use them and for whom they are most effective.

Acknowledgments

This work was supported by The Netherlands Cardiovascular Research Initiative, an initiative with support of the Dutch Heart Foundation, CVON2016-12 BENEFIT, ZonMw (The Netherlands Organization for Health Research and Development), Leiden University, and the members of the BENEFIT consortium. Furthermore, the authors would like to thank Fiona Brosig (Research Master student, Psychology, Leiden University) for her assistance in developing the intervention, recruiting participants, and running the experiment.

Authors' Contributions

DRDB, TR, and AWME helped in the study design; DRDB, SP, and TK performed the intervention and app design and development; DRDB, TRCR, and performed the data acquisition; DRDB, TR, AWME, and LDB performed the data analysis and interpretation; DRDB, TR, SP, TK, and AWME drafted the manuscript; and DRDB, TR, TRCR, SP, TK, SAL, LDB, VRJ, RAK, AWME, and THAB revised the manuscript. All the authors gave their final approval and agreed to be accountable for all aspects of the work, ensuring integrity and accuracy.

Conflicts of Interest

SP and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Institute for Implementation Science at the University of Zürich, the School of Medicine and Institute of Technology Management at the University of St. Gallen, and the Department of Management, Technology, and Economics at Eidgenössische Technische Hochschule Zürich. Centre for Digital Health Interventions is funded in part by CSS, a Swiss health insurer. TK is also the cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies were involved in the design, interpretation, and analysis of this study, or in writing this paper.

Multimedia Appendix 1
Flowchart.
[DOCX File , 2364 KB - jmir_v24i10e38339_app1.docx]

Multimedia Appendix 2
Baseline survey.
[DOCX File , 14 KB - jmir_v24i10e38339_app2.docx]

Multimedia Appendix 3
Final survey.
[DOCX File , 16 KB - jmir_v24i10e38339_app3.docx]

Multimedia Appendix 4
Main analysis total sample.
[DOCX File , 16 KB - jmir_v24i10e38339_app4.docx]

Multimedia Appendix 5
Sensitivity checks.
[DOCX File , 13 KB - jmir_v24i10e38339_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 432 KB - jmir_v24i10e38339_app6.pdf]

References

1. Woessner MN, Tacey A, Levinger-Limor A, Parker AG, Levinger P, Levinger I. The evolution of technology and physical inactivity: the good, the bad, and the way forward. Front Public Health 2021 May 28;9:655491 [FREE Full text] [doi: 10.3389/fpubh.2021.655491] [Medline: 34123989]
2. Exercise or Physical Activity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/fastats/exercise.htm [accessed 2021-11-04]
3. Anderson E, Durstine JL. Physical activity, exercise, and chronic diseases: a brief review. Sports Med Health Sci 2019 Dec;1(1):3-10 [FREE Full text] [doi: 10.1016/j.smhs.2019.08.006] [Medline: 35782456]
4. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec 25;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
5. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
6. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
7. Rhodes RE, de Bruijn G. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013 May 11;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
8. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
9. Laibson D. Golden eggs and hyperbolic discounting. Q J Econ 1997 May 01;112(2):443-478. [doi: 10.1162/003355397555253]
10. O'Donoghue T, Rabin M. Doing it now or later. Am Econ Rev 1999;89(1):103-124.
11. Wang Y, Sloan FA. Present bias and health. J Risk Uncertain 2018 Oct 25;57(2):177-198 [FREE Full text] [doi: 10.1007/s11166-018-9289-z] [Medline: 31244508]
12. Hunter RF, Tang J, Hutchinson G, Chilton S, Holmes D, Kee F. Association between time preference, present-bias and physical activity: implications for designing behavior change interventions. BMC Public Health 2018 Dec 19;18(1):1388 [FREE Full text] [doi: 10.1186/s12889-018-6305-9] [Medline: 30567532]
13. Kurti AN, Davis DR, Redner R, Jarvis BP, Zvorsky I, Keith DR, et al. A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change. Transl Issues Psychol Sci 2016 Jun;2(2):128-152 [FREE Full text] [doi: 10.1037/tps0000067] [Medline: 27777964]
14. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2020 Nov;54(21):1259-1268. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
15. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JP, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis. Prev Med 2015 Jun;75:75-85 [FREE Full text] [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]

16. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014 Mar 11;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
17. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
18. Lipman SA, Brouwer WB, Attema AE. A QALY loss is a QALY loss is a QALY loss: a note on independence of loss aversion from health states. Eur J Health Econ 2019 Apr 18;20(3):419-426 [FREE Full text] [doi: 10.1007/s10198-018-1008-9] [Medline: 30229374]
19. Halpern SD, French B, Small DS, Saulsgiver K, Harhay MO, Audrain-McGovern J, et al. Randomized trial of four financial-incentive programs for smoking cessation. N Engl J Med 2015 May 28;372(22):2108-2117 [FREE Full text] [doi: 10.1056/NEJMoa1414293] [Medline: 25970009]
20. Patel MS, Asch DA, Volpp KG. Framing financial incentives to increase physical activity among overweight and obese adults. Ann Intern Med 2016 Oct 18;165(8):600. [doi: 10.7326/l16-0280]
21. Stedman-Falls LM, Dallery J. Technology-based versus in-person deposit contract treatments for promoting physical activity. J Appl Behav Anal 2020 Sep 16;53(4):1904-1921. [doi: 10.1002/jaba.776] [Medline: 32939749]
22. The WayBetter way to manage weight. WayBetter. URL: https://www.waybetter.com [accessed 2022-02-22]
23. Ready to finally stickK to your Goals? stickK. URL: https://www.stickk.com [accessed 2022-02-22]
24. Burns RJ, Rothman AJ. Comparing types of financial incentives to promote walking: an experimental test. Appl Psychol Health Well Being 2018 Jul 19;10(2):193-214. [doi: 10.1111/aphw.12126] [Medline: 29671957]
25. Cawley J, Price JA. A case study of a workplace wellness program that offers financial incentives for weight loss. J Health Econ 2013 Sep;32(5):794-803. [doi: 10.1016/j.jhealeco.2013.04.005] [Medline: 23787373]
26. Sykes-Muskett BJ, Prestwich A, Lawton RJ, Armitage CJ. The utility of monetary contingency contracts for weight loss: a systematic review and meta-analysis. Health Psychol Rev 2015;9(4):434-451. [doi: 10.1080/17437199.2015.1030685] [Medline: 25933128]
27. Jarvis BP, Dallery J. Internet-based self-tailored deposit contracts to promote smoking reduction and abstinence. J Appl Behav Anal 2017 Apr;50(2):189-205 [FREE Full text] [doi: 10.1002/jaba.377] [Medline: 28211949]
28. Budworth L, Prestwich A, Sykes-Muskett B, Khatun K, Ireland J, Clancy F, et al. A feasibility study to assess the individual and combined effects of financial incentives and monetary contingency contracts on physical activity. Psychol Sport Exercise 2019 Sep;44:42-50. [doi: 10.1016/j.psychsport.2019.04.021]
29. Donlin Washington W, McMullen D, Devoto A. A matched deposit contract intervention to increase physical activity in underactive and sedentary adults. Transl Issues Psychol Sci 2016 Jun;2(2):101-115. [doi: 10.1037/tps0000069]
30. Krebs CA, Nyein KD. Increasing physical activity in adults using self-tailored deposit contacts. Behav Analysis Res Pract 2021 Aug;21(3):174-183. [doi: 10.1037/bar0000222]
31. Royer H, Stehr M, Sydnor J. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. Am Econ J Appl Econ 2015 Jul 01;7(3):51-84. [doi: 10.1257/app.20130327]
32. Adjerid I, Loewenstein G, Purta R, Striegel A. Gain-loss incentives and physical activity: the role of choice and wearable health tools. Manag Sci 2022 Apr 28;68(4):2642-2667. [doi: 10.1287/mnsc.2021.4004]
33. Lipman SA. One size fits all? Designing financial incentives tailored to individual economic preferences. Behav Public Policy 2020 Jun 22:1-15 [FREE Full text] [doi: 10.1017/bpp.2020.21]
34. Giné X, Karlan D, Zinman J. Put your money where your butt is: a commitment contract for smoking cessation. Am Econ J Applied Econ 2010 Oct 01;2(4):213-235. [doi: 10.1257/app.2.4.213]
35. Finkelstein EA, Tham K, Haaland BA, Sahasranaman A. Applying economic incentives to increase effectiveness of an outpatient weight loss program (TRIO) - a randomized controlled trial. Soc Sci Med 2017 Jul;185:63-70. [doi: 10.1016/j.socscimed.2017.05.030] [Medline: 28554160]
36. John LK, Loewenstein G, Troxel AB, Norton L, Fassbender JE, Volpp KG. Financial incentives for extended weight loss: a randomized, controlled trial. J Gen Intern Med 2011 Jun;26(6):621-626 [FREE Full text] [doi: 10.1007/s11606-010-1628-y] [Medline: 21249462]
37. Adams J, Giles EL, McColl E, Sniehotta FF. Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. Health Psychol Rev 2014;8(3):286-295. [doi: 10.1080/17437199.2013.848410] [Medline: 25053215]
38. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
39. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
40. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
41. de Buisonjé D. Less carrot more stick: deposit contracts for health behaviour change. Open Source Framework. 2021 Mar 31. URL: https://osf.io/34ygt [accessed 2022-09-23]

42. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: a novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: overview and preliminary evaluation in the public health context. In: Proceedings of the 2015 Wireless Telecommunications Symposium (WTS). 2015 Presented at: 2015 Wireless Telecommunications Symposium (WTS); Apr 15-17, 2015; New York, NY, USA. [doi: 10.1109/wts.2015.7117255]
43. Kowatsch T, Volland D, Shih C, Rüegger D, Künzler F, Barata F, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: Designing the Digital Transformation. Cham: Springer; 2017.
44. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
45. Tinschert P, Rassouli F, Barata F, Steurer-Stey C, Fleisch E, Puhan MA, et al. Prevalence of nocturnal cough in asthma and its potential as a marker for asthma control (MAC) in combination with sleep quality: protocol of a smartphone-based, multicentre, longitudinal observational study with two stages. BMJ Open 2019 Jan 07;9(1):e026323 [FREE Full text] [doi: 10.1136/bmjopen-2018-026323] [Medline: 30617104]
46. MobileCoach. URL: https://www.mobile-coach.eu/ [accessed 2022-09-23]
47. Google Fit. URL: https://www.google.com/fit/ [accessed 2022-09-23]
48. HealthKit. Apple Developer Inc. URL: https://developer.apple.com/documentation/healthkit [accessed 2022-09-23]
49. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
50. Polese JC, E Faria GS, Ribeiro-Samora GA, Lima LP, Coelho de Morais Faria CD, Scianni AA, et al. Google fit smartphone application or Gt3X Actigraph: which is better for detecting the stepping activity of individuals with stroke? A validity study. J Bodyw Mov Ther 2019 Jul;23(3):461-465. [doi: 10.1016/j.jbmt.2019.01.011] [Medline: 31563356]
51. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
52. Yao J, Tan CS, Lim N, Tan J, Chen C, Müller-Riemenschneider F. Number of daily measurements needed to estimate habitual step count levels using wrist-worn trackers and smartphones in 212,048 adults. Sci Rep 2021 May 05;11(1):9633 [FREE Full text] [doi: 10.1038/s41598-021-89141-3] [Medline: 33953288]
53. Togo F, Watanabe E, Park H, Shephard RJ, Aoyagi Y. Meteorology and the physical activity of the elderly: the Nakanojo Study. Int J Biometeorol 2005 Nov 26;50(2):83-89. [doi: 10.1007/s00484-005-0277-z] [Medline: 16044348]
54. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. Am J Health Promot 2006 Aug 25;21(2):85-89. [doi: 10.4278/0890-1171-21.2.85] [Medline: 17152246]
55. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Amsterdam, Netherlands: Elsevier Science; 1988.
56. Halpern SD, Asch DA, Volpp KG. Commitment contracts as a way to health. BMJ 2012 Jan 30;344(jan30 1):e522 [FREE Full text] [doi: 10.1136/bmj.e522] [Medline: 22290100]
57. Mukherjee S, Sahay A, Pammi V, Srinivasan N. Is loss-aversion magnitude-dependent? Measuring prospective affective judgments regarding gains and losses. Judgm Decis Mak 2017 Jan;12(1):81-89 [FREE Full text]
58. Ludolph R, Schulz PJ. Does regulatory fit lead to more effective health communication? A systematic review. Soc Sci Med 2015 Mar;128:142-150. [doi: 10.1016/j.socscimed.2015.01.021] [Medline: 25617673]
59. Finkelstein EA, Bilger M, Baid D. Effectiveness and cost-effectiveness of incentives as a tool for prevention of non-communicable diseases: a systematic review. Soc Sci Med 2019 Jul;232:340-350 [FREE Full text] [doi: 10.1016/j.socscimed.2019.05.018] [Medline: 31129504]
60. Tudor-Locke C, Leonardi C, Johnson WD, Katzmarzyk PT, Church TS. Accelerometer steps/day translation of moderate-to-vigorous activity. Prev Med 2011 Jul;53(1-2):31-33. [doi: 10.1016/j.ypmed.2011.01.014] [Medline: 21295063]
61. Gal D, Rucker DD. The loss of loss aversion: will it loom larger than its gain? J Consum Psychol 2018 Apr 20;28(3):497-516. [doi: 10.1002/jcpy.1047]

Abbreviations

ANCOVA: analysis of covariance

Edited by R Kukafka; submitted 29.03.22; peer-reviewed by M Kamada, J Zink; comments to author 22.06.22; revised version received 04.08.22; accepted 11.08.22; published 06.10.22.

Please cite as:
de Buisonjé DR, Reijnders T, Cohen Rodrigues TR, Prabhakaran S, Kowatsch T, Lipman SA, Bijmolt THA, Breeman LD, Janssen VR, Kraaijenhagen RA, Kemps HMC, Evers AWM
Investigating Rewards and Deposit Contract Financial Incentives for Physical Activity Behavior Change Using a Smartphone App: Randomized Controlled Trial
J Med Internet Res 2022;24(10):e38339
URL: https://www.jmir.org/2022/10/e38339
doi:10.2196/38339
PMID:36201384

Original Paper

Developing the "Choosing Health" Digital Weight Loss and Maintenance Intervention: Intervention Mapping Study

Iga Palacz-Poborczyk[1*], BSc, MA; Paulina Idziak[1*], BSc, MA; Anna Januszewicz[1], PhD; Aleksandra Luszczynska[1], PhD; Eleanor Quested[2,3], PhD; Felix Naughton[4], PhD; Martin S Hagger[5,6], PhD; Sherry Pagoto[7], PhD; Peter Verboon[8], PhD; Suzanne Robinson[9], PhD; Dominika Kwasnicka[1,10], MA, MSc, PhD

[1]Faculty of Psychology, SWPS University of Social Sciences and Humanities, Wroclaw, Poland

[2]Physical Activity and Well-being Research Group, enAble Research Institute, Curtin University, Perth, Australia

[3]Curtin School of Population Health, Curtin University, Perth, Australia

[4]Behavioural and Implementation Science Research Group, School of Health Sciences, University of East Anglia, Norwich, United Kingdom

[5]Psychological Sciences, University of California, Merced, CA, United States

[6]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[7]Department of Allied Health Sciences, The UConn Center for mHealth and Social Media, University of Connecticut, Storrs (Mansfield), CT, United States

[8]Department of Psychology and Educational Sciences, Open Universiteit Nederland, Heerlen, Netherlands

[9]Faculty of Health, Deakin University, Melbourne, Australia

[10]NHMRC Centre of Research Excellence in Digital Technology to Transform Chronic Disease Outcomes, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Dominika Kwasnicka, MA, MSc, PhD
Faculty of Psychology, SWPS University of Social Sciences and Humanities
Aleksandra Ostrowskiego 30b
Wroclaw, 53-238
Poland
Phone: 48 501205532
Email: dkwasnicka@gmail.com

Abstract

Background: Digital health promotion programs tailored to the individual are a potential cost-effective and scalable solution to enable self-management and provide support to people with excess body weight. However, solutions that are widely accessible, personalized, and theory- and evidence-based are still limited.

Objective: This study aimed to develop a digital behavior change program, *Choosing Health*, that could identify modifiable predictors of weight loss and maintenance for each individual and use these to provide tailored support.

Methods: We applied an Intervention Mapping protocol to design the program. This systematic approach to develop theory- and evidence-based health promotion programs consisted of 6 steps: development of a logic model of the problem, a model of change, intervention design and intervention production, the implementation plan, and the evaluation plan. The decisions made during the Intervention Mapping process were guided by theory, existing evidence, and our own research—including 4 focus groups (n=40), expert consultations (n=12), and interviews (n=11). The stakeholders included researchers, public representatives (including individuals with overweight and obesity), and experts from a variety of relevant backgrounds (including nutrition, physical activity, and the health care sector).

Results: Following a structured process, we developed a tailored intervention that has the potential to reduce excess body weight and support behavior changes in people with overweight and obesity. The *Choosing Health* intervention consists of tailored, personalized text messages and email support that correspond with theoretical domains potentially predictive of weight outcomes for each participant. The intervention content includes behavior change techniques to support motivation maintenance, self-regulation, habit formation, environmental restructuring, social support, and addressing physical and psychological resources.

Conclusions: The use of an Intervention Mapping protocol enabled the systematic development of the *Choosing Health* intervention and guided the implementation and evaluation of the program. Through the involvement of different stakeholders,

including representatives of the general public, we were able to map out program facilitators and barriers while increasing the ecological validity of the program to ensure that we build an intervention that is useful, user-friendly, and informative. We also summarized the lessons learned for the *Choosing Health* intervention development and for other health promotion programs.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-040183

(J Med Internet Res 2022;24(10):e34089) doi:10.2196/34089

KEYWORDS

behavior change; behavior maintenance; behavioral theory; weight loss; overweight; obesity; randomized controlled trial; digital health; within-person design; Intervention Mapping

Introduction

Background

Worldwide, overweight and obesity are a major public health concern showing continuous increase over the past 4 decades [1]. Excess body weight is a risk factor for multiple chronic health conditions and diseases, including cardiovascular disease, cancer, and type 2 diabetes [2]. Overweight and obesity are also risk factors for severe illness in patients with COVID-19 [3]. The development of programs that support people with overweight and obesity to lose and maintain weight loss is urgently needed [4], especially programs that promote healthy nutrition, physical activity, and health behavior change as a means to lose weight and maintain weight loss. Several evidence-based state-of-the-art weight loss programs exist [5,6]; however, often they are not tailored specifically to the individual. Personalization and tailoring of health promotion programs can increase the cost and complexity of the program [7]; however, it also has multiple advantages.

Intervention tailoring involves adapting the intervention based on specific characteristics of the recipient [8]. A recent systematic review of tailored digital health interventions for weight loss [9] showed that tailored interventions were generally more effective in supporting weight loss compared with generic interventions or waitlist controls. Information can be tailored to the participants in several ways that are reported to vary in effectiveness; most interventions apply *descriptive tailoring*, meaning that participants are provided with information that is tailored based on their responses to a series of questions [9]. For example, participants who report low levels of self-efficacy may receive intervention content that will support them in developing mastery to perform a specific task [10].

An alternative way to tailor the intervention is through *inferential tailoring*, meaning that participants are monitored over a period and information regarding their characteristics, behavioral predictors, and behavioral outcomes is collected (eg, by means of ecological momentary assessment [EMA]) [11,12]. Through the process of longitudinal data collection, the researcher gathers and then analyzes data about the participant (collected using digital technologies) that can provide inferences about what the strongest predictors of relevant outcomes are [13] and, therefore, how the content of the intervention can be most appropriately tailored for them. Inferential tailoring can also account for trends in data and be applied at the time and situation when the intervention is most desirable (eg, by means of just-in-time adaptive interventions [14,15]).

To the best of our knowledge, to date, only 1 study has explored the predictors and outcomes associated with weight loss maintenance in individuals using EMA and N-of-1 designs [16]. The results showed that individuals who lost weight had unique psychological profiles (ie, specific psychological predictors of health behavior change) that could be accounted for in subsequent interventions. However, this study did not use the collected data to personalize the interventions.

To develop personalized interventions, within-person studies exploring weight loss trajectories and changes in cognition and outcomes are needed [17]. Web-based interventions for weight loss and weight loss maintenance in people with overweight and obesity have small to moderate overall effects compared with minimal or control conditions [18]. To move the science forward toward personalized behavioral medicine, new data-driven methods of tailoring and digital health support need to be tested [19]. We now have the opportunity to develop health interventions that are truly individualized and tailored to individuals' psychological profiles by applying new technologies that support unobtrusive data collection and EMA.

Intervention mapping protocols [20] can be used as a powerful tool to develop effective personalized interventions in a systematic manner. Intervention mapping [21] is a comprehensive framework that can be used to develop new interventions and health promotion programs, adjust existing programs to new contexts and realities [20], and develop implementation strategies [22]. It is an ecological approach that includes active involvement of stakeholders in program development and follows a series of 6 established steps. The steps undertaken are sequential; however, they are also iterative, and intervention developers often move back and forth between the steps to design the most optimal intervention.

Intervention mapping protocols have been successfully applied to design several health promotion programs, including interventions to decrease sedentary behavior [23], improve self-management of type 2 diabetes [24], improve self-care in heart failure [25], prevent risks and hazards in occupational settings [26], and in several other health-related settings. The approach was also applied to design programs that aimed to tackle overweight and obesity in various populations, including children [27,28], pregnant women [29], adolescents [30], adults [31], and workers [32]. Most of the aforementioned health promotion programs used digital technologies; however, none of them included a health behavior change program that was tailored to each individual and based on the participants' own data via *inferential tailoring*.

Objectives

In this manuscript, we describe the systematic development of the *Choosing Health* program. An Intervention Mapping protocol guided decisions regarding program objectives, behavior change methods, production, implementation, and evaluation. All decisions made during the Intervention Mapping process were guided by theory [33], evidence [16,34], and our own research undertaken during the intervention design phase (including focus groups, local expert consultations, and interviews). The *Choosing Health* program is a complex health promotion intervention that applies digital technology to collect participants' EMA data and then use them to provide a tailored intervention. The intervention is aimed at supporting individuals to change their physical activity and nutritional behaviors to ultimately help them lose weight and maintain weight loss. This study aimed to develop a digital behavior change program, *Choosing Health*, following a comprehensive Intervention Mapping protocol.

Methods

Study Design

This was an Intervention Mapping study; all study materials, standard operating procedures, and design decisions were documented, and the intervention content was published in the Open Science Framework repository [35]. This Intervention Mapping study resulted in the development of the *Choosing Health* program, which is currently being evaluated through a randomized controlled trial (RCT); this trial was registered at ClinicalTrials.gov (NCT04291482), and the trial protocol was published elsewhere [36].

The Intervention Mapping procedure included 6 steps that comprised several tasks integrating theory and evidence [20]. The completion of all the steps served as a blueprint for designing, implementing, and evaluating the *Choosing Health* intervention based on theoretical, empirical, and practical information. The 6 steps and related tasks of the Intervention Mapping process are described in the following sections and summarized in Multimedia Appendix 1 [9,18,33,36-38].

Step 1: Needs Assessment

In step 1, we established the planning group (ie, study authors) and conducted a needs assessment to create a logic model of the health problem (Multimedia Appendix 2). The intervention planning group guided the design, implementation, and evaluation of the *Choosing Health* program. This group consisted of 11 key stakeholders representing a variety of backgrounds supported by public representatives, including those representing the program target group (ie, people with overweight and obesity), nutrition and physical activity experts, health care practitioners, and implementers of the program (ie, individuals who could support future rollout of the program). The characteristics of various stakeholders are further described in the Step 4: Intervention Production section. The planning group assessed the issue of overweight and obesity in Poland [39], relevant behaviors, environmental factors, and their associated changeable determinants in the population (ie, individuals with overweight and obesity who require relevant support to lose weight and maintain weight loss). By means of a scoping review, we researched and described overweight and obesity and their impact on quality of life, stating environmental and behavioral changeable determinants. We researched and described the context of the intervention, including contextual factors, community and setting, and defined program goals.

Step 2: Identifying Objectives

In step 2, we created a logic model of change (Multimedia Appendix 3) defining specific program outcomes and objectives. Following the Intervention Mapping protocol, the planning group worked collaboratively to create a matrix that combined performance objectives (rows) and relevant changeable determinants (columns), listing in the cells specific change objectives. For each change objective, we specified who and what would change as a result of the proposed intervention, setting the foundation for the *Choosing Health* intervention.

Step 3: Intervention Design

Step 3 consisted of generating theory-based program themes, components, scopes, and sequences. The planning group chose theories that were relevant to the program and decided to include 5 theoretical themes from a recent theory review of behavior change maintenance as underpinning the program [33]. On the basis of the change objectives and determinants identified in the logic model, we also selected change methods and behavior change techniques (BCTs) underpinning the program (which are described in the protocol [36]). Theoretical themes were then mapped to specific BCTs [40], which are also described in the trial protocol [36]. For instance, the theoretical theme of habit formation was supported by specific BCTs such as performing the same behavior in the same context and adding cues to the environment (eg, setting reminders to exercise and putting fresh fruit and vegetables in the lunch pack) so that the context elicits the behavior. We also selected methods of delivery of the program (ie, practical applications to deliver change methods). The planning group chose digital health delivery via text messages, emails, and an intervention book (or e-book) to ensure that the proposed intervention was scalable if proven effective.

Step 4: Intervention Production

The aim of step 4 was to develop and refine the program structure and organization, preparation of program materials—including drafting theory- and evidence-based messages (emails, text messages, and book)—and program protocols. During this stage, we pretested and refined the materials through focus groups, expert ratings, and interviews.

Focus Groups

We conducted 4 focus groups (framed as user engagement workshops, of which 3/4 (75%) were conducted face-to-face and 1/4 (25%) were held on the web because of the COVID-19 pandemic) between November 2019 and May 2020. The focus group participants were recruited through Facebook (event advertisements posted on health-related pages) and through websites listing local events. They were also advertised through newsletters (of the university and of local health-related organizations) and posters placed in community venues and the university. Representatives of the general population (n=40),

including some people with overweight (10/40, 25%; 8/10, 80% women and 2/10, 20% men) and obesity (6/40, 15%; 4/6, 67% women and 2/6, 33% men), took part in the focus groups to discuss the project's rationale, aim, proposed format, and materials. The focus group participants' mean age was 31.55 (SD 13.15, range 19-65) years, and 22% (9/40) men and 78% (31/40) women took part, with most having a high school education (21/40, 52%) and some having higher education (11/40, 28% Bachelor of Arts or Bachelor of Science and 8/40, 20% Master of Arts or Master of Science. Their average BMI was 24.09 (SD 6.26, range 16.94-34.09; 1/40, 2% of the participants specified their height but not their weight).

The focus group participants assessed and rated the sample intervention materials to assess their clarity (on a scale of 1=unclear to 10=clear), attractiveness (on a scale of 1=unattractive to 10=attractive), and informativeness (on a scale of 1=uninformative to 10=informative). They rated some of the emails (25/96, 25%) and text messages (340/757, 44.9%). Materials were discussed within the group, and the pros and cons were elaborated on. The focus groups were audio recorded, transcribed (by PI), and verified (by IPP), and the transcripts were analyzed verbatim using the framework method [41] in NVivo software (version 12; QSR International) [42]. After familiarization with the transcripts, the first coder (IPP) generated initial themes and indexed the codes in a preliminary framework. These codes and preliminary themes were discussed with the second coder (PI), who independently coded 50% of the transcripts and provided feedback on the themes. The final set of themes was generated using an iterative approach, and all disagreements were discussed with a third researcher (DK) until a consensus was reached.

Expert Rating

The full set of intervention materials, including 109 emails and 759 text messages, was pretested with psychology, physical activity, and nutrition experts (n=12). The experts were recruited through the researchers' network as well as through web-based message boards and Facebook groups for professionals with relevant expertise. The experts were Polish, based in Poland, and they reviewed materials written in Polish. The experts had a mean age of 30.42 (SD 10.9, range 24-64) years and were 8% (1/12) men and 92% (11/12) women; 33% (4/12) had an MSc in Nutrition, 33% (4/12) had an MSc in Psychology, 25% (3/12) had an MSc in Public Health, and 8% (1/12) had a PhD in Health Sciences (including physical activity background). All text messages were assessed by at least two experts who rated the content using the same measures as the ones used during the focus groups to assess content attractiveness and informativeness and, in addition, emotional reactions (*How did it make you feel?* on a scale of 1 indicating negative reactions to 10 indicating positive reactions).

The experts were asked to assign each text message to relevant theoretical domains, with clear definitions of each domain provided (eg, habits, stress, and obstacles). The experts were asked to indicate their first, second, and third choice for the domain that the message aligned with. The experts did not have to have any background in behavioral science to assign messages to theoretical domains as clear definitions were provided and examples were given. They also provided additional open-ended comments if they had any feedback or reflections regarding specific text messages or emails. The experts completed this task in a Microsoft Excel form in their own time. The theoretical domains and definitions of the theoretical constructs were based on a comprehensive theory review [33].

Interviews

We also asked 11 representatives of the general population (n=6, 55% men and n=5, 45% women; mean age 39.27, SD 16.32, range 18-72 years) to evaluate the intervention book or e-book (our program participants had a choice between a physical book and an e-book). Interviewed participants were recruited through the researchers' networks. Each person read through the whole book and, by means of unstructured interviews (conducted by IPP), provided feedback on content, comprehensibility, user-friendliness of the design, and inclusiveness. The key points from each interview were summarized and noted by the interviewer, and the book was revised in line with the suggestions given.

The materials (emails, text messages, and book) were iteratively revised by 4 members of the project team (IPP, PI, DK, and AJ) and continuously adapted based on insights from the focus groups, interviews, and study experts. During the intervention content development stage (June 2020-August 2020), the core team (IPP, PI, DK, and AJ) met 11 times; each meeting took 2 to 3 hours, approximately 30 hours in total. The *Choosing Health* program protocol, including frequency, intensity, and sequence, was also discussed and agreed upon. All materials were developed in Polish and later translated into English and published on the Open Science Framework website. All study measures were forward and backward translated [43] if language-specific versions of the questionnaires were not available. All questions were adjusted for culture- and language-specific appropriateness and piloted with Polish speakers (n=15), with changes made in line with the feedback received. The project team members (IPP, PI, AJ, and DK) met 3 times (approximately 12 hours in total) to finalize the translation and adaptation of the questions and measurement tools for the trial (October 2019-November 2019).

Step 5: Implementation Plan

In step 5, we defined the intervention adaptation, implementation, and sustainability plan developing matrices defining change objectives to promote *Choosing Health* program adaptation and use. These objectives were operationalized forming theory-informed plans for intervention adaptation and implementation [44,45]. Through discussion, the planning group identified potential program users (adopters, implementers, and maintainers) considering both the initial program test (RCT) and if the program was to be widely implemented. Behavioral outcomes were defined and linked to the behavioral and environmental determinants. The resulting change objectives for program use were used to map the intervention for potential adopters, implementers, and maintainers designing the intervention implementation plan, which is further described in the Results section.

Step 6: Evaluation Plan

In the final step, we planned how to best evaluate the program effects, costs, and processes. A specific evaluation plan was developed by the core planning group, and the trial protocol was published [36]. We defined the mechanisms of intervention effectiveness informed by the previous Intervention Mapping steps. Following the Intervention Mapping protocol [20], we listed the effect, cost, and process evaluation research questions that are listed in the protocol [36]. We developed indicators and measures of success by defining study measures, measurement points, and thresholds for effectiveness based on the previous literature [17]. The intervention is currently ongoing through an RCT with an embedded N-of-1 study and ongoing cost and process evaluations.

Ethics Approval

Ethics approval was granted by the Faculty of Psychology, SWPS University of Social Sciences and Humanities, Poland (approval 03/P/12/2019).

Results

In this study, we used the Intervention Mapping approach following the aforementioned steps (Multimedia Appendix 1). In step 1, a needs assessment was used to define the problem—namely, high levels of obesity and overweight in Poland (reaching >53.3%) [39,46] and the need to design effective, cost-effective, and scalable programs that can support people in weight loss and subsequent weight loss maintenance. The impact on quality of life was prominent, with people with overweight and obesity reporting lower physical and mental health [39,46]. Several environmental and behavioral determinants were described and listed, including limited access to weight loss programs, the high cost of weight loss programs, obstacles to accessing healthy foods (eg, perceived as more expensive than unhealthy foods), and obstacles to engaging in physical activity (eg, perceived lack of time and limited access). Contextual factors included personal, family, work, and broader community influences, which could both enable and hinder behaviors conducive to weight loss and weight loss maintenance. The variety of determinants and contexts that needed to be considered pointed toward the need for a highly personalized and cost-effective program.

In step 2, the program's objectives were specified—namely, to develop a program that could support self-guided personalized weight loss, including behavior changes in physical activity, nutrition, and prompting psychological changes (in motivation, habits, self-regulation, resources, and context). The list of combined performance objectives and relevant changeable determinants included the following: individuals who complete the program will need to complete 2 key phases to lose weight and maintain weight loss. First, we need to learn about their individual predictors of weight loss and weight loss maintenance (therefore, we will encourage program users to self-monitor their determinants—theory-based constructs including motivation, habits, self-regulation, resources, and context). Subsequently, we will intervene on the strongest predictors of behavioral outcomes, providing relevant intervention content. For each determinant, we mapped the corresponding theoretical explanations and techniques [36]. We predicted that there are several changeable determinants that are relevant to each program user; however, each user is likely to have a different profile of determinants that are the most predictive of weight change and maintenance.

In step 3, techniques fitting the problem and objectives were chosen. On the basis of theory and evidence, we divided our intervention into 5 conceptual domains (maintained motivation, habit, self-regulation, resources, and environmental influences) and, within these domains, suitable BCTs were identified [36]. For instance, to support habit formation, we prompted rehearsal and repetition of the behavior in the same context so that the context elicited the behavior. We mapped out BCTs to each domain and operationalized them in intervention materials, including text messages, emails, and e-book.

In step 4, we conducted focus groups to refine the intervention content. A total of 40 participants took part in the focus groups, and the key themes discussed were analyzed and divided into 2 groups of themes: intervention content and form of program delivery. The first theme had three main subthemes: (1) the participant being an active agent in the change process, (2) inclusivity of the information provided, and (3) problem-solving. The second theme also had three main subthemes: (1) ensuring that the content was informative, (2) unambiguity of the provided information, and (3) including direct actionable messages. In Table 1, we include *lessons learnt* from the focus groups in relation to intervention content and form and direct quotes from focus group participants that align with different themes and subthemes.

The focus group participants' mean scores for the proposed text message content were relatively high on a scale of 1 to 10 (mean 8.38; clarity mean 9.27, SD 1.32; attractiveness mean 8.48, SD 1.24; informativeness mean 8.6, SD 1.33); higher scores reflected positive results, and lower scores reflected negative results for each category. These findings were corroborated in the focus group discussions (Table 1).

Experts rated the quality of the text messages as moderately high (mean 7.21; positive emotions mean 6.95, SD 1.41; attractiveness mean 7.23, SD 1.38; informativeness mean 7.47, SD 1.61). All text messages rated below an average of 4.5 across all categories were excluded or adjusted, and text messages that did not fit specific themes were reallocated or adjusted. In total, we excluded 3.7% (28/759) of text messages and 0.9% (1/109) of emails that were considered inappropriate or scored low overall.

In step 5, we identified potential program users as adults with overweight and obesity living in Poland. Initial program users were individuals living in Wroclaw and nearby areas as the initial test of the program (via RCT) required face-to-face assessments to objectively measure weight. Implementers of the program initially included the researchers involved in the program development and research assistants. Future program implementers and maintainers (if the program is proven effective) could include representatives from the government, health care representatives recommending the program, and community representatives. One of the routes to intervention implementation that we are assessing now is wide

implementation through the program partner Lifestyle Medicine [47,48]—an organization that provides medical training for health care practitioners educating them on the principles of behavior change and advocating to promote health behavior change in patients and minimize the overmedicalization of people with overweight and obesity. If the intervention is effective, it could contribute to lowering overweight and obesity rates in Poland, resulting in health improvements and cost savings.

In step 6, we generated a plan for cost, effect, and process evaluations. Currently, the program is being evaluated through an RCT assessing between-group effects (intervention vs control), and it is also being evaluated within people looking at the trajectories of change investigated through EMA using an N-of-1 design and inferential tailoring [36]. The resulting program is evidence-based, delivered through technology (text messages, email, and book), and tailored to each participant based on the data gathered through EMA. The evaluation plan follows the principles of process evaluation defined by the Medical Research Council (United Kingdom) following the guidelines for developing and evaluating complex interventions [37].

Table 1. Lessons learnt for the *Choosing Health* program and for other intervention developers from the focus groups undertaken (N=40)[a].

Theme and theme description	Lessons learned	Example quotes
Intervention content		
The participant being an active agent in the change process: study participants much preferred the messages that treated them as experts in their own behavior change. Any messages that could come across as condescending or coming from the perspective of a "teacher" or a person who "knows it all" or "knows better" were considered inappropriate.	• Study participants need to be treated as equal partners in the behavior change and behavior maintenance process. • Understanding of personal needs and preferences is key to providing useful intervention content. • Each message needs to contain elements of flexibility (the participant may want to take the suggestion on or not; they do not need to follow the suggestions fully).	• Condescending, stereotypical, and negative messages were unacceptable: "Not everyone who carries extra kilograms sits nonstop in front of the TV and eats crisps. We can't speak to them [intervention users] as if they did not have a clue that a week without the TV or a week without crisps is possible. The worse thing we could do is to look down on them." [Participant 14, woman, aged 24 years, BMI 29] • Intervention aims should be personalized and defined with the study participants: "I would simply ask what are the intentions of this person, what exactly motivates them? Why are they taking part in your program? Probably they want to lose weight but you need to understand other factors too..." [Participant 7, woman, aged 30 years, BMI 20] • To many participants, the provided information was not new and often complemented what they already knew and what they had already experienced: "From my own experience I can say that the feeling of hunger is just so personal. I had to relearn to understand when I'm hungry, when I'm full and when I've totally overeaten. Since childhood I was 'trained' to eat like a horse, to just feel more than full. I had to relearn to eat till I'm almost full, so I feel slightly unsatisfied. Some people still need to learn it and work on it." [Participant 9, man, aged 47 years, BMI 20] • People's levels of motivation and motivation sources vary, and interventions need to account for that: "Social support is very important but if other people don't want to support me, they should at least not criticise my choices. I look for support or at least lack of criticism of what I do. Maybe other people find it helpful to be criticised, for me, I find it really demotivating." [Participant 9, man, aged 47 years, BMI 20] • All messages suggesting that physical activity needs to be chosen in line with personal preferences were rated positively: "I really get on board with this, I really like that you suggest that physical activity doesn't need to be forced, and that I can just pick whatever I like, as long as I am active." [Participant 39, man, aged 39 years, BMI 25] • Most messages need to give the participant some options and choices. The participants prefer to choose what fits their lifestyle and preferences: "I love this message—I like that you say that there is not one type of food that makes someone feel better—one person may like nuts, other one may prefer fish, I just really like how you pointed out that this is all personal." [Participant 38, man, aged 65 years, BMI 30]

Theme and theme description	Lessons learned	Example quotes
Inclusivity of the information provided: it was important to tailor messages so that they fit in with people who are of different socioeconomic statuses or different personal circumstances, prefer different leisure activities, and have different health statuses and professions.	• The person's identity needs to be considered when defining intervention content. • The information provided needs to be inclusive, especially when discussing social support.	• Participants who did not have close family or lived far away from their family felt excluded when reading the messages pointing toward fun social family activities: "Someone who I don't actually even know, writes to me and says hug a family member, and I'm alone, I do not have any close family, I would get so p***** off, and sorry to phrase it like that, but I would just not continue with this." [Participant 32, woman, aged 44 years, BMI 28]
Problem-solving: study participants wanted to receive positive messages that motivated them to problem solve. The superficial approach of "it's all good" and "you can do it!" was not perceived as helpful. The participants needed some acknowledgment that weight loss is not easy and often comes with barriers and difficulties.	• People usually knew what the negative consequences were, and they did not need to be reminded. • They needed constructive suggestions for how to best problem solve. • Messages based on fear and negative emotions were considered unhelpful.	• Participants appreciated messages that emphasized their psychological resources and constructive ways of using self-motivation: "The message I really, really like is this one: 'Think about the day when you decided to join Choosing Health program! What motivated you to join? Note down thoughts that you had then.' I really liked this message coz people often undertake challenges and then half way through they forget why they actually doing it. The motivation is gone, and sometimes it's enough to just remind someone why are they doing it. Remembering your past success, can really reinforce your motivation and help you look more positively towards the future." [Participant 24, woman, aged 24 years, BMI 26] • The messages that described unpleasant situations, evoked negative emotions, and reminded the participants of some negative past events but did not include any actionable solutions that needed to be avoided: "Imagine, I'm in a good mood, having a really good day, everything going well and then I'm getting one of your messages, this one, it says 'consider what's causing stress in your life and think about how you could tackle it and change it'—So now what? I'm doing my exercise, drinking water, I eat healthily and now what? I'm stressing thinking oh dear God...my husband, all the debt I have..." [Participant 29, woman, aged 34 years, BMI 28] • The participants rated positively the messages that encouraged them to self-monitor and pointed them toward the strategies that they could implement immediately to improve: "I really like these messages that said that I should write down certain things, note what motivates me, and note what my goals are. That was great, a systematic way of doing things, if I write it down, I will remember it. If I read your text and I'm on the go, I may remember it but I may not..." [Participant 30, woman, aged 37 years, BMI 21]
Intervention form		

Theme and theme description	Lessons learned	Example quotes
Ensuring that the content is informative: the participants really appreciated the fact that the intervention was evidence-based. The expectations were high in terms of providing the most recent psychological knowledge. The participants wanted to receive fresh and novel content, and they wanted to develop their own knowledge.	• Just the fact that participants receive messages may be motivating, but this motivation is not long-lived if the content is not informative. • To maintain participants' engagement, they need to receive evidence-based, state-of-the-art, engaging, and original content. • People do not want to be overloaded with the information.	• It is important that the participants learn something new that they did not know before: "I really liked it because the information provided was interesting and some of it was surprising for me, especially when you elaborated on different causes of stress. It was clear, it made me feel good, I simply learnt new information." [Participant 36, woman, aged 25 years, BMI 17] • People want to read evidence-based information: "Some of the messages were just too simple. I don't want to sound big-headed but for me that was just way too simple. I would add (to the emails) at least few lines describing some background evidence, at least something showing that it's actually based on scientific evidence." [Participant 26, man, aged 24 years, BMI 20] • The intervention should encourage them to learn more: "Maybe you could encourage people to be more conscious and to learn more: 'Check if what you think is healthy, is actually healthy and good for you?' [...] A while ago I went to the shops, I had some time and I saw minced meat. I usually don't buy that type of meat, and then I read what's on the label, and I couldn't believe it! [...] Maybe you could consider that people need to seek to educate themselves a bit more." [Participant 21, woman, aged 34 years, BMI 34] • The intervention should include valuable messages without sounding superficial: "I just have a general suggestion—for me messages that include phrases 'healthy food,' 'balanced diet,' 'balance' sound a bit superficial, as people don't really know what that means." [Participant 26, man, aged 24 years, BMI 20] • Explaining psychological and physiological mechanisms related to weight loss was always welcome, especially if communicated in a clear way (but only in emails as SMS text messages were perceived as too short to clearly communicate the meaning and dependencies): "I was positively surprised to read the email 'What stress actually is?' [...] This message had a lot of important, easy to digest info, stress is so common these days so I was glad to read more on this topic." [Participant 36, woman, aged 25 years, BMI 17] • Psychological evidence was expected, and the participants wanted to learn more about the mechanisms of action and behavior change techniques: "So where are all the psychological aspect here? You say monitor eating to not overdo it, but how am I meant to do that?!" [Participant 32, woman, aged 44 years, BMI 28]
Unambiguity of the provided information: messages should have a clear meaning and preferably cover only 1 topic at a time.	• When the message includes multiple topics, it is more difficult to understand, reflect on, and implement.	• The participants did not perceive it helpful when healthy eating, physical activity, education, and social support were all mentioned in one message: "For me it is problematic that you talk about healthy eating and physical activity. Someone may be concentrating hard on eating healthily but not really on improving physical activity. [...] I would probably just emphasize one or the other as tackling both at the same time is hard." [Participant 7, woman, aged 30 years, BMI 20]

Theme and theme description	Lessons learned	Example quotes
Including direct actionable messages: participants did not appreciate idioms or references to literature, culture, or pop culture. The use of humor was controversial and had very diverse reception.	• Study participants preferred direct, clear, and actionable messages.	• Participants did not appreciate "mental shortcuts," and every change of topic in the message had to be clearly announced to them: "This message about feeling grateful—I completely don't get what's the relationship between feeling grateful and losing weight or improving health." [Participant 38, man, aged 65 years, BMI 30] • Idioms and messages including humor were negatively received. Losing weight and maintaining weight loss are often perceived as sensitive topics, and the use of humor is often considered inappropriate: "...again you are using an idiom here—and I'm fairly sure that not everyone is able to understand it in this context." [Participant 40, man, aged 65 years, BMI 32]

[a]Interview data were analyzed with the aim of improving intervention content and form (ie, look and feel, intensity, and sequence) so that the main themes were predefined before the analysis process. The subthemes emerged from the discussions and data analysis.

Discussion

Principal Findings

The overall aim of this Intervention Mapping study was to inform the *Choosing Health* weight loss program by applying existing theory, evidence, and principles of public engagement throughout the planning process. The Intervention Mapping protocol steps were closely followed to ensure that the intervention was useful, user-friendly, and designed with the users to ensure the clarity, attractiveness, and positive reception of the proposed program. The focus groups, expert consultations, and interviews showed that the program content and proposed format were rated highly, and the elements that did not meet a specific threshold (4.5 out of 10) were adjusted in line with feedback or omitted. Active involvement of individuals from the target group of the *Choosing Health* intervention—namely, people with overweight or obesity—enabled the specification of the key needs and wants of the recipients of the intervention. The materials were produced iteratively and sequentially, and the mode of delivery was thoroughly discussed with the potential users to ensure the feasibility of the proposed program.

In relation to the intervention content, the main results were that program participants need to be actively involved in the change process, which aligns with theory [49] and previous interventions [50]. The information provided needs to be inclusive and encourage the participants to actively problem solve while they are changing their behavior and maintaining it in the long term [51]. In terms of the format of the program, the key results were that it needs to be informative and unambiguous and include direct and actionable messages, aligning with other recommendations for the development of health behavior change programs [52,53]. Previous studies that also gathered EMA data on daily predictors of weight loss [16] did not use inferential tailoring to provide health behavior change advice and information. This will be the first study that uses longitudinal data to then provide tailored support.

This study has several strengths. The key strength is the use of the thorough and rigorous Intervention Mapping protocol that served as a tool and provided us with vocabulary to comprehensively map out and plan the proposed intervention [20]. Designing interventions iteratively with the users and using a variety of study methods ensures that the interventions have high ecological validity [54]. The intervention was designed by the core group (the study authors) in close collaboration with public representatives and field experts (eg, nutritionists and physical activity experts) to increase the ecological validity of the proposed program. Engaging the target audience in the intervention design ensures that the programs are suitable and useful and that they target relevant determinants [55]; it also ensures that we account for diversity in the participant population [56]. The intervention content was designed to be tailored to specific theoretical domains that were predictive of effects (that will be assessed by participants through EMA to define the strongest predictors of outcomes). We consulted the experts and asked them to allocate each element of the intervention content to a specific theoretical domain to ensure that the content fits the theoretical domains. This validation ensured that we targeted the correct determinants that are the strongest predictors of outcomes for each individual.

The study limitations include lack of involvement of some key stakeholders in the planning group. Namely, representative policy-makers from the local or national government and IT did not participate in the planning group. To ensure scalability and long-term maintenance of the proposed program, it would need to be integrated with existing intervention programs or policies operating within the health care system or local communities [57,58]. The planning group met with representatives from the Ministry of Health of Poland, who initially expressed support to promote the project and implement it at a national scale if proven effective through the nationwide health care and health promotion website [59]. However, following structural and personnel changes in the national government, the plan was no longer feasible to implement. Liaising with the Ministry of Health during the COVID-19 pandemic also proved difficult. Other studies and health promotion programs emphasize how valuable it is to engage policy-makers in the intervention development and implementation processes [57,58], and we are hoping to meaningfully engage with them in the future. To scale the

intervention (if effective) and make it accessible to people in Poland, a national body (government and health care sector) needs to endorse the intervention and embed it within the existing support structures. The COVID-19 pandemic has emphasized the need for self-guided remote support to improve health.

The proposed intervention has specific components that combine different technology aspects—data harvesting via EMA, automated text messaging, automated emails, and book; therefore, this intervention could further benefit from the active involvement of technology developers, data scientists, and computer programmers. The researchers working on the project designed the technology interface using existing components (eg, automated messaging systems). However, to further enhance the scalability of the intervention, the engagement of computer scientists would allow us to implement more sophisticated data analytics and intervention setup methods. Future interventions need to include automated machine learning algorithms that would allow for the analysis of data in real time and automated setup of the intervention to improve efficiencies and reduce the resources needed [60,61]. Machine learning is a valuable and increasingly necessary tool for health promotion and for the modern health care system [62], and it should be applied in future personalized interventions.

In Poland, the prevalence of overweight and obesity is higher in men than in women (46.8% vs 32.2% for overweight, respectively, and 20.1% vs 18.1% for obesity, respectively); however, the sample recruited for the focus groups comprised predominantly women (31/40, 78%) and a predominantly normal weight BMI category (24/40, 60%). In the ideal scenario, most of the focus group participants would have been men, and most or all participants would have been overweight and obese. Specific challenges, including social stigma and stereotypes associated with dieting and weight loss programs, played a role in recruiting a more representative sample of the user population. However, we have explored whether there are any differences among the opinions and feedback given by men and women and also by people who fall into BMI categories below and above 25, and we have not found any pronounced differences.

The key take-home messages from our *Choosing Health* Intervention Mapping study were (1) involving several types of stakeholders as early as possible in the Intervention Mapping process, (2) iterating the intervention with various groups of stakeholders and learning from the incoming evaluation data, and (3) allowing for flexibility in health promotion programs. As the intervention was designed to be delivered on the web, the COVID-19 pandemic did not have an impact on the delivery of the intervention; however, it has affected study data collection that was initially intended to be conducted face-to-face. In addition, one of the focus groups had to be conducted on the web. Several research teams working worldwide are facing similar challenges, and specific technology solutions are being developed to support these teams in data collection during the pandemic [54,55]. Currently, developing health promotion programs that can be fully delivered on the web is important and needed.

Conclusions

We developed a comprehensive weight loss and maintenance intervention targeting important behavioral and contextual determinants. The development of the intervention followed comprehensive steps of the Intervention Mapping process and was grounded in theory and relevant literature. Future evaluation studies will investigate the program effectiveness, cost-effectiveness, and process and further analyze the relevance and utility of the specific program components. The findings from this study may be particularly useful for other intervention developers who are also planning to design and implement personalized digital health weight loss interventions targeting behavioral nutrition, physical activity, and health behavior change.

Acknowledgments

The *Choosing Health* project was carried out within the HOMING program of the Foundation for Polish Science, cofinanced by the European Union under the European Regional Development Fund (grant POIR.04.04.00-00-5CF3/18-00; HOMING 5/2018). The authors would like to acknowledge the study participants and expert volunteers for their time and effort in attending the co-design workshops and providing feedback on the proposed study procedures and materials.

Data Availability

The data sets generated and analyzed during this study are available in the Open Science Framework repository [35].

Authors' Contributions

DK, EQ, MSH, and FN conceived the project and obtained project funding. All authors (IPP, PI, DK, AL, MSH, EQ, SP, PV, SR, AJ, and FN) have made conceptual contributions to the project design and procedures. PV is a trial statistician who designed a data analysis plan together with FN and DK. SR is a trial health economist who designed an economic evaluation plan. SP and AJ provided practitioner insights. PI and IPP managed the day-to-day activities of the trial and executed the study. DK is the project lead. IPP, PI, and DK drafted the manuscript. All authors read, edited, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 6 steps of Intervention Mapping undertaken during the Choosing Health program development combining methodologies and results.
[DOCX File , 17 KB - jmir_v24i10e34089_app1.docx]

Multimedia Appendix 2
A logic model of the health problem.
[PNG File , 248 KB - jmir_v24i10e34089_app2.png]

Multimedia Appendix 3
The logic model of change.
[PNG File , 134 KB - jmir_v24i10e34089_app3.png]

References

1. Obesity and overweight. World Health Organization. 2021 Jun 9. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-08-01]
2. Swinburn BA, Kraak VI, Allender S, Atkins VJ, Baker PI, Bogard JR, et al. The global syndemic of obesity, undernutrition, and climate change: the Lancet commission report. Lancet 2019 Feb 23;393(10173):791-846. [doi: 10.1016/S0140-6736(18)32822-8] [Medline: 30700377]
3. Wang J, Zhu L, Liu L, Zhao X, Zhang Z, Xue L, et al. Overweight and obesity are risk factors of severe illness in patients with COVID-19. Obesity (Silver Spring) 2020 Nov 23;28(11):2049-2055 [FREE Full text] [doi: 10.1002/oby.22979] [Medline: 32735706]
4. Fastenau J, Kolotkin RL, Fujioka K, Alba M, Canovatchel W, Traina S. A call to action to inform patient-centred approaches to obesity management: development of a disease-illness model. Clin Obes 2019 Jun 11;9(3):e12309 [FREE Full text] [doi: 10.1111/cob.12309] [Medline: 30977293]
5. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018 Feb 10;391(10120):541-551. [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
6. Hunt K, Wyke S, Gray CM, Anderson AS, Brady A, Bunn C, et al. A gender-sensitised weight loss and healthy living programme for overweight and obese men delivered by Scottish Premier League football clubs (FFIT): a pragmatic randomised controlled trial. Lancet 2014 Apr;383(9924):1211-1221. [doi: 10.1016/s0140-6736(13)62420-4]
7. Lairson DR, Chung TH, Huang D, Stump TE, Monahan PO, Christy SM, et al. Economic evaluation of tailored web versus tailored telephone-based interventions to increase colorectal cancer screening among women. Cancer Prev Res (Phila) 2020 Mar;13(3):309-316 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-19-0376] [Medline: 31969343]
8. Rakowski W. The potential variances of tailoring in health behavior interventions. Ann Behav Med 1999 Dec;21(4):284-289. [doi: 10.1007/bf02895959]
9. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
10. Vandelanotte C, Short CE, Plotnikoff RC, Rebar A, Alley S, Schoeppe S, et al. Are web-based personally tailored physical activity videos more effective than personally tailored text-based interventions? Results from the three-arm randomised controlled TaylorActive trial. Br J Sports Med 2021 Mar 03;55(6):336-343. [doi: 10.1136/bjsports-2020-102521] [Medline: 33144346]
11. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavorial medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
12. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
13. Hobbs N, Dixon D, Johnston M, Howie K. Can the theory of planned behaviour predict the physical activity behaviour of individuals? Psychol Health 2013 Mar;28(3):234-249. [doi: 10.1080/08870446.2012.716838] [Medline: 22943555]
14. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014 Jul;13(3):13-17. [doi: 10.1109/MPRV.2014.46]
15. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S(Suppl):1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
16. Kwasnicka D, Dombrowski SU, White M, Sniehotta FF. N-of-1 study of weight loss maintenance assessing predictors of physical activity, adherence to weight loss plan and weight change. Psychol Health 2017 Jun 21;32(6):686-708. [doi: 10.1080/08870446.2017.1293057] [Medline: 28323457]

17. Kwasnicka D, Naughton F. N-of-1 methods: a practical guide to exploring trajectories of behaviour change and designing precision behaviour change interventions. Psychol Sport Exercise 2020 Mar;47:101570. [doi: 10.1016/j.psychsport.2019.101570]
18. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
19. Kwasnicka D, Inauen J, Nieuwenboom W, Nurmi J, Schneider A, Short CE, et al. Challenges and solutions for N-of-1 design studies in health psychology. Health Psychol Rev 2019 Jun 23;13(2):163-178. [doi: 10.1080/17437199.2018.1564627] [Medline: 30626274]
20. Bartholomew-Eldredge L, Markham C, Ruiter R, Fernández M, Kok G, Parcel G. Planning Health Promotion Programs An Intervention Mapping Approach. Hoboken, New Jersey, United States: Wiley; 2016.
21. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
22. Fernandez ME, Ten Hoor GA, van Lieshout S, Rodriguez SA, Beidas RS, Parcel G, et al. Implementation mapping: using intervention mapping to develop implementation strategies. Front Public Health 2019 Jun 18;7:158 [FREE Full text] [doi: 10.3389/fpubh.2019.00158] [Medline: 31275915]
23. Berninger N, Ten Hoor GA, Plasqui G, Kok G, Peters G, Ruiter R. Sedentary work in desk-dominated environments: a data-driven intervention using intervention mapping. JMIR Form Res 2020 Jul 20;4(7):e14951 [FREE Full text] [doi: 10.2196/14951] [Medline: 32706695]
24. 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]
25. Athilingam P, Clochesy JM, Labrador MA. Intervention mapping approach in the design of an interactive mobile health application to improve self-care in heart failure. Comput Inform Nurs 2018 Mar;36(2):90-97. [doi: 10.1097/CIN.0000000000000383] [Medline: 28901967]
26. Bakhuys Roozeboom MC, Wiezer N, Boot C, Bongers P, Schelvis R. Use of intervention mapping for occupational risk prevention and health promotion: a systematic review of literature. Int J Environ Res Public Health 2021 Feb 11;18(4):1775 [FREE Full text] [doi: 10.3390/ijerph18041775] [Medline: 33670376]
27. Ball GD, Mushquash AR, Keaschuk RA, Ambler KA, Newton AS. Using Intervention Mapping to develop the Parents as Agents of Change (PAC) intervention for managing pediatric obesity. BMC Res Notes 2017 Jan 13;10(1):43 [FREE Full text] [doi: 10.1186/s13104-016-2361-3] [Medline: 28086848]
28. Pittson H, Wallace L. Using intervention mapping to develop a family-based childhood weight management programme. J Health Serv Res Policy 2011 Apr;16 Suppl 1:2-7. [doi: 10.1258/jhsrp.2010.010076] [Medline: 21460343]
29. Merkx A, Ausems M, de Vries R, Nieuwenhuijze MJ. Come On! Using intervention mapping to help healthy pregnant women achieve healthy weight gain. Public Health Nutr 2017 Mar 15;20(9):1666-1680. [doi: 10.1017/s1368980017000271]
30. Singh AS, Chin A Paw MJ, Kremers SP, Visscher TL, Brug J, van Mechelen W. Design of the Dutch Obesity Intervention in Teenagers (NRG-DOiT): systematic development, implementation and evaluation of a school-based intervention aimed at the prevention of excessive weight gain in adolescents. BMC Public Health 2006 Dec 16;6(1):304 [FREE Full text] [doi: 10.1186/1471-2458-6-304] [Medline: 17173701]
31. van Beurden SB, Greaves C, Abraham C, Lawrence NS, Smith J. ImpulsePal: the systematic development of a smartphone app to manage food temptations using intervention mapping. Digit Health 2021;7:20552076211057667 [FREE Full text] [doi: 10.1177/20552076211057667] [Medline: 34868619]
32. Kwak L, Kremers S, Werkman A, Visscher T, van Baak MA, Brug J. The NHF-NRG In Balance-project: the application of Intervention Mapping in the development, implementation and evaluation of weight gain prevention at the worksite. Obesity Rev 2007 Jul;8(4):347-361. [doi: 10.1111/j.1467-789x.2006.00304.x]
33. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
34. Kwasnicka D, Dombrowski SU, White M, Sniehotta FF. 'It's not a diet, it's a lifestyle': a longitudinal, data-prompted interview study of weight loss maintenance. Psychol Health 2019 Aug 23;34(8):963-982. [doi: 10.1080/08870446.2019.1579913] [Medline: 30905184]
35. A theory-based digital intervention to promote weight loss and weight loss maintenance (Choosing Health). OSF Home. 2020 Apr 3. URL: https://osf.io/sf264/ [accessed 2022-09-19]
36. Kwasnicka D, Luszczynska A, Hagger MS, Quested E, Pagoto SL, Verboon P, et al. Theory-based digital intervention to promote weight loss and weight loss maintenance (Choosing Health): protocol for a randomised controlled trial. BMJ Open 2020 Nov 23;10(11):e040183 [FREE Full text] [doi: 10.1136/bmjopen-2020-040183] [Medline: 33234638]
37. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]

38. Dombrowski SU, Avenell A, Sniehott FF. Behavioural interventions for obese adults with additional risk factors for morbidity: systematic review of effects on behaviour, weight and disease risk factors. Obes Facts 2010 Dec;3(6):377-396 [FREE Full text] [doi: 10.1159/000323076] [Medline: 21196792]
39. Gańczak M, Miazgowski T, Kożybska M, Kotwas A, Korzeń M, Rudnicki B, et al. Changes in disease burden in Poland between 1990-2017 in comparison with other Central European countries: a systematic analysis for the Global Burden of Disease Study 2017. PLoS One 2020;15(3):e0226766 [FREE Full text] [doi: 10.1371/journal.pone.0226766] [Medline: 32119685]
40. 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]
41. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
42. Zamawe F. The implication of using NVivo software in qualitative data analysis: evidence-based reflections. Malawi Med J 2015 Mar 24;27(1):13-15 [FREE Full text] [doi: 10.4314/mmj.v27i1.4] [Medline: 26137192]
43. Degroot A, Dannenburg L, Vanhell J. Forward and backward word translation by bilinguals. J Memory Language 1994 Oct;33(5):600-629. [doi: 10.1006/jmla.1994.1029]
44. Baumann A, Cabassa L, Stirman S. Adaptation in dissemination and implementation science. In: Dissemination and Implementation Research in Health: Translating Science to Practice (2nd edn) Dissemination and Implementation Research in Health: Translating Science to Practice (2nd edition). New York: Oxford University Press; 2017.
45. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
46. Noncommunicable diseases country profiles Poland. World Health Organization. URL: https://www.who.int/nmh/countries/pol_en.pdf [accessed 2021-08-02]
47. Ezinwa N. Global lifestyle medicine: for people who need it the most but have it the least. Am J Lifestyle Med 2020;14(5):541-545 [FREE Full text] [doi: 10.1177/1559827620912721] [Medline: 32922239]
48. Mondala M, Sannidhi D. Catalysts for change: accelerating the lifestyle medicine movement through professionals in training. Am J Lifestyle Med 2019;13(5):487-494 [FREE Full text] [doi: 10.1177/1559827619844505] [Medline: 31523214]
49. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
50. Kwasnicka D, Ntoumanis N, Hunt K, Gray CM, Newton RU, Gucciardi DF, et al. A gender-sensitised weight-loss and healthy living program for men with overweight and obesity in Australian Football League settings (Aussie-FIT): a pilot randomised controlled trial. PLoS Med 2020 Aug 6;17(8):e1003136 [FREE Full text] [doi: 10.1371/journal.pmed.1003136] [Medline: 32760144]
51. Presseau J, Sniehotta FF, Francis JJ, Campbell NC. Multiple goals and time constraints: perceived impact on physicians' performance of evidence-based behaviours. Implement Sci 2009 Nov 26;4(1):77 [FREE Full text] [doi: 10.1186/1748-5908-4-77] [Medline: 19941655]
52. Kwasnicka D, Vandelanotte C, Rebar A, Gardner B, Short C, Duncan M, et al. Comparing motivational, self-regulatory and habitual processes in a computer-tailored physical activity intervention in hospital employees - protocol for the PATHS randomised controlled trial. BMC Public Health 2017 May 26;17(1):518 [FREE Full text] [doi: 10.1186/s12889-017-4415-4] [Medline: 28549476]
53. Quested E, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Kerr DA, Hunt K, et al. Protocol for a gender-sensitised weight loss and healthy living programme for overweight and obese men delivered in Australian football league settings (Aussie-FIT): a feasibility and pilot randomised controlled trial. BMJ Open 2018 Oct 17;8(10):e022663 [FREE Full text] [doi: 10.1136/bmjopen-2018-022663] [Medline: 30337315]
54. Kwasnicka D, Ten Hoor GA, Hekler E, Hagger MS, Kok G. Proposing a new approach to funding behavioural interventions using iterative methods. Psychol Health 2021 Jul 27;36(7):787-791. [doi: 10.1080/08870446.2021.1945061] [Medline: 34180340]
55. Staley K. 'Is it worth doing?' Measuring the impact of patient and public involvement in research. Res Involv Engagem 2015 Jul 31;1(1):6 [FREE Full text] [doi: 10.1186/s40900-015-0008-5] [Medline: 29062495]
56. Oliver S, Liabo K, Stewart R, Rees R. Public involvement in research: making sense of the diversity. J Health Serv Res Policy 2015 Jan;20(1):45-51. [doi: 10.1177/1355819614551848] [Medline: 25228453]
57. Campbell D, Redman S, Jorm L, Cooke M, Zwi A, Rychetnik L. Increasing the use of evidence in health policy: practice and views of policy makers and researchers. Aust New Zealand Health Policy 2009 Aug 24;6:21 [FREE Full text] [doi: 10.1186/1743-8462-6-21] [Medline: 19698186]

58. Guieu A, Zhang W, Lafort Y, Decat P, De Meyer S, Wang S, et al. Practical lessons for bringing policy-makers on board in sexual and reproductive health research. BMC Health Serv Res 2016 Nov 11;16(1):649 [FREE Full text] [doi: 10.1186/s12913-016-1889-1] [Medline: 27835973]
59. Serwis Ministerstwa Zdrowia i Narodowego Funduszu Zdrowia. Strona główna | Pacjent. URL: https://pacjent.gov.pl/ [accessed 2022-09-22]
60. Conroy D, Lagoa C, Hekler E, Rivera D. Engineering person-specific behavioral interventions to promote physical activity. Exerc Sport Sci Rev 2020 Oct;48(4):170-179 [FREE Full text] [doi: 10.1249/JES.0000000000000232] [Medline: 32658043]
61. Gonul S, Namli T, Huisman S, Laleci Erturkmen GB, Toroslu I, Cosar A. An expandable approach for design and personalization of digital, just-in-time adaptive interventions. J Am Med Inform Assoc 2019 Mar 01;26(3):198-210 [FREE Full text] [doi: 10.1093/jamia/ocy160] [Medline: 30590757]
62. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018 Apr 03;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]

Abbreviations

BCT: behavior change technique
EMA: ecological momentary assessment
RCT: randomized controlled trial

Edited by R Kukafka; submitted 05.10.21; peer-reviewed by AR Mohseni, M allman-Farinelli, J Fiedler, H Mehdizadeh, I Wilson, K Champion, L McGowan; comments to author 16.12.21; revised version received 10.05.22; accepted 29.05.22; published 18.10.22.

Please cite as:
Palacz-Poborczyk I, Idziak P, Januszewicz A, Luszczynska A, Quested E, Naughton F, Hagger MS, Pagoto S, Verboon P, Robinson S, Kwasnicka D
Developing the "Choosing Health" Digital Weight Loss and Maintenance Intervention: Intervention Mapping Study
J Med Internet Res 2022;24(10):e34089
URL: https://www.jmir.org/2022/10/e34089
doi:10.2196/34089
PMID:36256827

Original Paper

The Impact of a Theory-Based mHealth Intervention on Disease Knowledge, Self-efficacy, and Exercise Adherence Among Ankylosing Spondylitis Patients: Randomized Controlled Trial

Yuqing Song[1], MSN; Elizabeth Reifsnider[2], PhD; Yanling Chen[3], MSN; Ying Wang[3], MSN; Hong Chen[1], PhD

[1]West China School of Nursing, West China Hospital, Sichuan University, Chengdu, China

[2]Edson College of Nursing and Health Innovation, Arizona State University, Phoenix, AZ, United States

[3]Department of Rheumatology and Immunology, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Hong Chen, PhD
West China School of Nursing
West China Hospital
Sichuan University
No 37 Guoxue Xiang
Wuhou District
Chengdu, 610041
China
Phone: 86 189 8060 1733
Email: 1366109878@qq.com

Abstract

Background: Patient education is recommended as an integral part of disease management in ankylosing spondylitis (AS), a chronic rheumatic disease that predominantly affects young males and requires long-term disease management. Convenient and cost-effective approaches to deliver patient education are required to these patients.

Objective: This study aimed to examine the effects of a theory-based educational intervention delivered through a social networking app, WeChat, on disease knowledge, self-efficacy, exercise adherence, and health outcomes in Chinese AS patients.

Methods: This study was a single-blind randomized controlled trial conducted in a tertiary hospital in Chengdu, China. Eligible participants were randomly allocated to the intervention or control group. Participants in the control group received standard care. The intervention group received the health belief model (HBM)-based educational intervention, consisting of 4 individual educational sessions and educational information sharing through WeChat, the predominant social networking app in China. The primary outcomes were disease knowledge, self-efficacy, and exercise adherence. The secondary outcomes were disease activity and physical function. Data were collected at baseline and at the end of the intervention (12th week). Chi-square test, *t* test, Mann-Whitney *U* tests were used to examine the effects of educational intervention.

Results: This study included 118 patients with AS. The majority of participants were male (93/118, 78.8%). Around half of them were married (56/118, 47.5%), never smoked (70/118, 59.3%), and had college educational level or above (62/118, 52.5%). At posttest, participants in the intervention group had higher disease knowledge (all $P<.001$) and self-efficacy ($P<.001$), and a larger proportion of participants in the intervention group adhered to regular exercise routines than those in the control group ($P=.003$). The within-group analyses for the intervention group showed increases in all scores of disease knowledge (all $P<.001$) and self-efficacy score ($P<.001$), but only correct answer score ($P=.04$) and general knowledge score ($P=.002$) of disease knowledge in the control group improved. The within-group analysis for the control group found a decline of physical function ($P=.002$) but no significant change in disease activity ($P>.05$). The within-group analysis for the intervention group showed no significant change in disease activity or physical function ($P>.05$). At posttest, no statistically significant difference was found on disease activity or physical function between the intervention and control groups ($P>.05$).

Conclusions: The HBM-based educational intervention through WeChat can effectively improve patient disease knowledge, self-efficacy, and exercise adherence. WeChat is feasible and effective to deliver patient education for patients with chronic diseases such as AS. This mHealth intervention can be integrated into routine rheumatology care.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IPR-16009293; https://tinyurl.com/swxt8xk7

(J Med Internet Res 2022;24(10):e38501) doi:10.2196/38501

KEYWORDS

rheumatic disease; health belief model; mobile health; patient education; WeChat

Introduction

Background

Ankylosing spondylitis (AS) is a chronic rheumatic disease characterized by inflammatory back pain, morning stiffness, and reduced physical function that requires long-term management [1]. Regular disease management, including pharmacological therapy and exercise, is essential to control disease and prevent progression of the disease [1,2]. The inflammatory back pain and morning stiffness of AS patients improves with exercise but not with rest or inactivity [3]. Evidence reveals that exercise (such as stretching or aerobic exercise) can improve health outcomes, including pain, morning stiffness, and physical function, and the effects of exercise depend on patient adherence [4,5]. Exercise adherence refers to the extent to which people undertake the prescribed exercise from health care providers [5]. However, AS patients face severe challenges during disease management, such as lack of knowledge about the disease and nonadherence to medication and exercise [6-8]. Lack of knowledge about AS is a barrier to exercise and medication adherence [8]. Lack of exercise adherence negatively influences health outcomes among AS patients [8].

Patient education comprises educational activities designed to influence patient knowledge and health behaviors, enable patients to manage their disease, and optimize health outcomes [2,9]. Ndosi et al [10] revealed that patient education should aim at improving patient self-efficacy since self-efficacy is a predictor of health behaviors and health outcomes. Self-efficacy is defined as an individual's confidence in performing a specific behavior [11]. Previous studies indicated that patient education can increase disease knowledge, self-efficacy, and adherence in arthritis patients [10,12,13]. However, only several published studies explored the effects of patient education among AS patients [14-20]. These interventions mostly reported small sample sizes, the lack of theoretical basis, and found limited effects on self-efficacy and adherence, and inconsistent results on health outcomes (eg, disease activity, physical function) [15-20]. The intervention delivery methods of many previous studies relied heavily on face-to-face interactions, which can be difficult due to travel restrictions, time constraints due to any number of factors, and costs of missing work [18,21].

Mobile health (mHealth) interventions can deliver timely health service and overcome the obstacles of time, distance, and cost [22]. The wide use of mobile phones has increased the possibility of delivering through health-focused interventions via apps [23]. In recent years, WeChat has been the most popular social networking app in China, with over 1 billion monthly active users [24]. WeChat can offer free message, voice/video calls, and enhance effective communication and information sharing [23,25]. WeChat has been used as a tool for educational interventions in patients with cancer, hypertension, coronary heart disease, and these studies reported positive effects [23,26-29]. Evidence revealed that AS patients need available education through phones and apps [30]. However, there is little evidence of an educational intervention through WeChat for AS patients.

AS patients may benefit from effective theory-based interventions to improve health behaviors and health outcomes (eg, disease activity, physical function). The health belief model (HBM) is developed to explain how to change health behaviors and focuses on an individual's likelihood of engaging in healthy behaviors [11,31]. Previous studies revealed that the HBM is effective in developing interventions to change an individual's beliefs and healthy behaviors in cancer, pulmonary tuberculosis patients [32,33]. However, this theory has not been used to develop an educational intervention for AS patients.

Objectives

Based on these findings, a randomized controlled trial for AS patients was conducted and aimed to compare patient outcomes in a theory-based mHealth intervention via WeChat with standard care. Results regarding quality of life, depression, and selected clinical outcomes have been published elsewhere [34]. This paper describes the primary outcomes of this intervention, including disease knowledge, self-efficacy, exercise adherence, disease activity, and physical function. We hypothesized that the HBM-based mHealth intervention would improve the disease knowledge, self-efficacy, exercise adherence, physical function, and control disease activity of AS patients.

Theoretical Framework

The key constructs of HBM consist of perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action and self-efficacy [11,35]. In HBM, health behaviors are based on people's perceptions of susceptibility to and severity of health problems, barriers and benefits to enacting health behaviors and cues to action [31]. Self-efficacy can improve the efficacy of the model, and change subsequent health behaviors [11]. This intervention helped patients understand the severity of AS and provided strategies for managing their disease in order to improve self-efficacy, healthy behaviors, and achieve better health outcomes. We applied the theory to the AS intervention by using the constructs to guide the design of the intervention and match the intervention elements to the hypothesized outcomes. The primary goals of patient education are to transfer knowledge about disease and improve health behaviors [9]. Exercise adherence is a crucial health behavior related to AS patients, and self-efficacy is a predictor of health behaviors. This educational intervention was designed to explore short-term effects. Thus, we selected disease knowledge, self-efficacy, and exercise adherence as primary outcomes while health outcomes (ie, disease activity and physical function) were secondary outcomes.

Methods

Study Design

This study was a single-blinded randomized controlled trial conducted from March to December 2017. Written informed consents were obtained from all participants (legal guardians of participants under 18 years provided written informed consent) after they had received information about the study protocol.

Sample Size Calculation

The sample size was determined by self-efficacy score based on our pilot trial, in which standard deviation was 1.47 and mean difference between the two groups was 0.86. With a power of 0.80 and α=.05 (2-sided), each group required 47 participants. The final sample size was 114, allowing a 20% dropout.

Ethics Approval

The study was registered at the Chinese Clinical Trial Registry [ChiCTR-IPR-16009293], conducted in accordance with the Declaration of Helsinki, and approved by West China Hospital Medical Ethics Committee (ID: 20160364).

Participants and Recruitment

The Department of Rheumatology and Immunology from a tertiary hospital in a large city, serving a population of Southwest China, was the site of recruitment. Potential participants were recruited via convenience sampling during their routine care. Participants were included if they (1) were diagnosed with AS according to the Modified New York Classification Criteria for AS, (2) were aged 14 years or older (participants under 18 years need written informed consent signed by legal guardians), (3) could speak and understand Chinese, (4) could use WeChat and had a WeChat account, and (5) were willing to participate in this randomized controlled trial study. We excluded participants if they (1) had severe cognitive or mental problems (comprehension or expression problems, using psychotropic drugs), (2) had other rheumatic diseases, or (3) were participating in other research programs.

Randomization

To ensure participant assignment was truly randomized without human bias, a well-trained study coordinator generated a random allocation table in Excel (version 2010, IBM Corp). Before recruiting participants, the study coordinator placed randomized numbers in sealed opaque envelopes. Participants selected an envelope at random. Based on the selected number, participants were allocated to the intervention or control group. The researcher who collected the data and study coordinator who generated the random allocation table were blinded to group allocation.

Intervention

Control Group

Patients in the control group received only standard care, including basic health advice appropriate for AS patients. Standard care was provided to the participants (including legal guardians of participants under 18 years) after they were recruited and completed the baseline assessment at the Department of Rheumatology and Immunology. Basic health advice was given in person by a nurse, and paper handouts were provided to the participants for their own education.

Intervention Group

The intervention group received the 12-week theory-based educational intervention delivered by WeChat plus standard care. The research team developed the educational intervention based on the HBM, a literature review of other social media–based interventions, expert consultation, and a pilot study. Finally, we identified the core content and corresponding HBM construct: basic knowledge of AS (perceived severity of disease, perceived susceptibility), medication (perceived benefit of preventive action), exercise (perceived benefit of preventive action), daily life management (perceived benefit of preventive action, cues to action), psychological support (perceived barriers to preventive action, supporting perceived self-efficacy), and self-assessment (perceived barriers to preventive action, cues to action; Table 1).

In the baseline assessment, we added participants as friends in WeChat and taught them how to use WeChat. The intervention consisted of 2 parts: online individual education sessions and educational information sharing. The first part included 4 individual educational sessions via WeChat video/voice calls on the 2nd, 4th, 8th, and 12th week. Each session was conducted for 20 to 30 minutes. Researchers contacted participants via telephone calls if they were absent from WeChat intervention sessions for 3 times. During the calls, research nurses built trusting relationships with participants, exploring their needs, problems managing disease, and psychological concerns. The nurses then used storytelling to illustrate potential severity. Nurse coaching during the calls helped the patients to establish cues to action to exercise, take medications, and promote health behaviors. The nurses used nurse coaching, verbal persuasion, and peer experience to address mood changes and support efforts toward self-efficacy. They taught participants how to assess their health conditions using validated instruments. They also encouraged self-efficacy by highlighting positive changes and helping patients manage self-doubt when lapses occurred. During the individual educational sessions, the nurses assessed participant knowledge about AS, problems, and health behaviors (eg, taking medication, exercising) related to AS, so the nurses could ensure whether the intervention positively influenced the target themes (eg, basic knowledge, medication) and provide targeted education. The second part consisted of selected pictures, videos, and articles on the WeChat public account about the core content of the intervention. The nurses sent links to online information once a week to participants. Moreover, participants could chat with the nurses at any time when they encountered problems with disease management.

Table 1. Content of the theory-based mobile health intervention.

Theme	Content	HBM[a] construct	Method
Basic knowledge	Causes, pathogenesis, clinical symptoms, treatment, prognosis of AS[b]	• Perceived severity of disease and its impact on future • Perceived susceptibility to increasing limited mobility	• One-to-one WeChat call: using storytelling to illustrate potential severity • Linking to online information about AS basic knowledge
Medication	Treatment goals, importance of taking medication, medication management at home, side effect management, how to use a reminder for medication taking	• Perceived benefits of preventive action to maintain current status	• One-to-one WeChat call: using nurse coaching to highlight the benefit of medication taking • Linking to online information about medication adherence
Exercise	Benefits of exercise, exercise type and intensity, helping reduce the obstacles to exercise, how to exercise at home	• Perceived benefits of preventive action	• One-to-one WeChat call: using nurse coaching to highlight the benefit of exercise regularly • Online video • Linking to online information on how to exercise
Daily life management	Physical posture, sleep instruction, diet, joint protection, quit smoking, etc	• Perceived benefits of preventive action • Creating cues to action	• One-to-one WeChat call: using verbal persuasion and nurse coaching to highlight perceived benefits of healthy behaviors • Linking to online information on daily life management
Psychological support	Psychological management strategies, providing patients with psychological support	• Supporting perceived self-efficacy to manage disease • Overcoming perceived barriers to preventive actions	• One-to-one WeChat call: using nurse coaching, verbal persuasion, and peer experience to support efforts toward self-efficacy
Self-assessment	Teaching patients how to assess disease activity, function, and psychological status, etc	• Creating cues to action • Overcoming perceived barriers to preventive action	• One-to-one WeChat call • Sending online information on validated instruments

[a]HBM: health belief model.

[b]AS: ankylosing spondylitis.

Measures

Demographic Information

Participant demographic data included age, gender, marital status, educational level, income, medical insurance, smoking status, and disease duration.

Primary Outcomes

Primary outcomes included disease knowledge, self-efficacy, and exercise adherence. Disease knowledge refers to the level of knowledge about AS in patients with AS [36]. In this study, patients' level of knowledge of AS was assessed by the Assessment of Knowledge in Ankylosing Spondylitis Patients [37]. This instrument is divided into 4 areas: (1) general knowledge, etiology, symptoms, blood tests; (2) B27 antigen and inheritance; (3) drug treatment and physical therapy; and (4) joint protection, pacing, and priorities. The instrument has 14 questions with 72 potential answers, but only 25 answers are correct. The correct answer score (maximum possible=25) is obtained by giving 1 point to each correct answer, and the correct item score (maximum possible=14) is obtained by giving 1 point to each question with all the correct answers [36]. Higher scores indicate higher levels of knowledge about AS. Cronbach alpha of this instrument was .729 in this study.

Self-efficacy is defined as an individual's confidence in performing a specific behavior [11]. In this study, self-efficacy was measured using the Arthritis Self-Efficacy Scale–8 (ASES-8) [38]. ASES-8 included 2 items for pain subscales, 4 items from other symptoms subscales, and 2 items that related to keeping pain and fatigue from interfering with things the patients want to do [39]. The final score is 0 to 10, with higher scores indicating higher self-efficacy. ASES-8 had good reliability, validity, and adaptability in arthritis patients [38,40]. Cronbach alpha of ASES-8 in this study is .913.

Exercise adherence refers to the extent to which people undertake the exercise prescribed by health care providers [5]. In this study, adherence to exercise was examined using a self-reported statement as used in previous studies [16,41]: frequency of exercise per week (0, occasionally, 1-2 days, 3-4 days, 5-6 days, daily).

Secondary Outcomes

Disease activity and physical function reflect the main aspects of health outcomes among AS patients. Disease activity was measured by the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) [42], a patient-reported scale to assess the severity of major symptoms (fatigue, spinal pain, joint pain and swelling, areas of localized tenderness, and morning stiffness) in AS patients. Physical function was measured by the Bath

Ankylosing Spondylitis Functional Index (BASFI) [43], a patient-reported scale to assess patient function (eg, bending, reaching, changing position) and the ability to cope with everyday life. The final BASDAI and BASFI scores ranged from 0 to 10, with higher scores indicating higher disease activity and worse physical function; Cronbach α=.740 (BASDAI) and α=.956 (BASFI).

Data Collection

Data were collected at baseline and the 12th week. If participants had difficulty in reading or writing, the researcher would help them complete questionnaires. Baseline data were collected from participants and medical records. The posttest data were collected from participants when they came to the rheumatology clinic for routine care or through an online survey platform [44] or through telephone/WeChat call.

Statistical Analyses

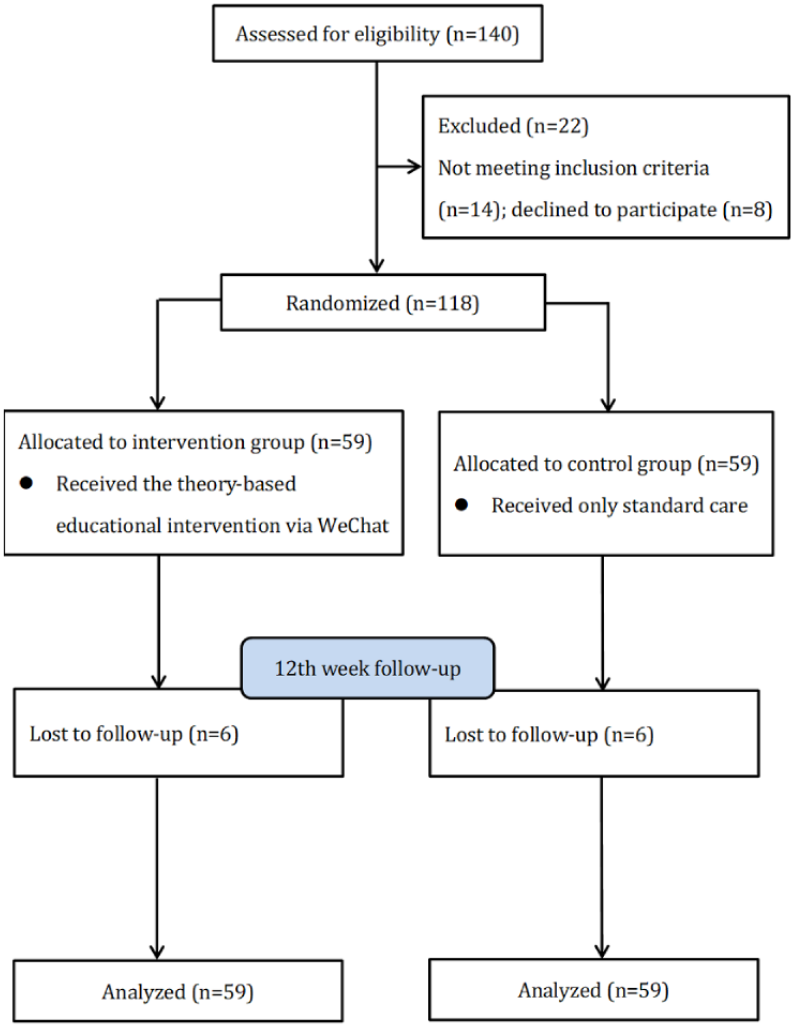
Data were analyzed using SPSS (version 22.0, IBM Corp) software. An intention-to-treat principle was used for analyses, and the last observation carried forward method was used for missing data assessment. Data were described as mean and standard deviation, median and interquartile range, and frequency and percentage. Independent sample *t* test, Mann-Whitney *U* test, and chi-square test were used to compare data between the intervention and control groups. Paired sample *t* tests were used to analyze the changes in outcomes from baseline to the 12th week within each group in continuous variables. $P<.05$ was considered statistically significant.

Results

Participants

A total of 118 participants were included and randomly allocated into the intervention (n=59) or control group (n=59). A total of 89.8% (106/118) of participants completed the study. Additionally, we included 118 participants in data analyses because the intention-to-treat principle was used. Figure 1 shows the study flowchart.

Figure 1. Flow diagram of the study.



Baseline Characteristics of Participants

The average age of participants was 29.9 years. The majority of participants were male (93/118, 78.8%). Around half of participants were married (56/118, 47.5%), never smoked (70/118, 59.3%), and had college educational level or above (62/118, 52.5%). There was no statistically significant difference in the variables between the 2 groups (Table 2).

Table 2. Baseline characteristics of participants in the intervention and control group.

Characteristic	Total (n=118)	Intervention (n=59)	Control (n=59)	$t^a/\chi^{2b}/Z^c$	*P* value
Age (years), mean (SD)	29.9 (8.23)	30.8 (8.82)	29.1 (7.58)	1.142[a]	.26
Gender, n (%)	—[d]	—	—	0.457[b]	.50
Male	93 (78.8)	45 (76.3)	48 (81.4)	—	—
Female	25 (21.2)	14 (23.7)	11 (18.6)	—	—
Educational level, n (%)	—	—	—	0.218[b]	.90
Junior high school or below	30 (25.4)	15 (25.4)	15 (25.4)	—	—
Senior high school	26 (22.0)	12 (20.3)	14 (23.7)	—	—
College or above	62 (52.5)	32 (54.2)	30 (50.8)	—	—
Marital status, n (%)	—	—	—	0[b]	>.99
Single/divorced	62 (52.5)	31 (52.5)	31 (52.5)	—	—
Married	56 (47.5)	28 (47.5)	28 (47.5)	—	—
Monthly per capita income (￥), n (%)	—	—	—	1.720[b]	.63
<2200	34 (28.8)	16 (27.1)	18 (30.5)	—	—
2200-3300	33 (28.0)	15 (25.4)	18 (30.5)	—	—
3300-5500	23 (19.5)	11 (18.6)	12 (20.3)	—	—
>5500	28 (23.7)	17 (28.8)	11 (18.6)	—	—
Medical insurance, n (%)	—	—	—	0.160[b]	.69
Self-pay	82 (69.5)	40 (67.8)	42 (71.2)	—	—
Medical insurance	36 (30.5)	19 (32.2)	17 (28.8)	—	—
Smoking status, n (%)	—	—	—	1.502[b]	.47
Current smoking	36 (30.5)	19 (32.2)	17 (28.8)	—	—
Never smoking	70 (59.3)	36 (61.0)	34 (57.6)	—	—
Quit smoking	12 (10.2)	4 (6.8)	8 (13.6)	—	—
Symptom duration, median (IQR)	5.00 (6.25)	6.00 (7.00)	5.00 (7.00)	–0.346[c]	.73
Diagnosis duration, median (IQR)	3.00 (6.00)	3.00 (6.00)	3.00 (6.00)	–0.577[c]	.56
Knowledge of AS[e]					
Correct item score, mean (SD)	6.41 (3.08)	6.64 (2.92)	6.17 (3.24)	0.836[a]	.41
Correct answer score, mean (SD)	15.29 (4.91)	15.61 (4.58)	14.97 (5.24)	0.711[a]	.48
General knowledge, mean (SD)	5.21 (1.86)	5.19 (1.82)	5.24 (1.91)	–0.148[a]	.88
B27 antigen and inheritance, median (IQR)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	–0.938[c]	.35
Drug treatment and physical therapy, mean (SD)	6.05 (2.16)	6.34 (1.91)	5.76 (2.37)	1.456[a]	.15
Joint protection, pacing and priorities, mean (SD)	3.22 (1.35)	3.34 (1.17)	3.10 (1.52)	0.952[a]	.34
Self-efficacy, mean (SD)	6.29 (1.93)	6.40 (1.91)	6.18 (1.97)	0.624[a]	.53
Disease activity, mean (SD)	3.28 (1.89)	3.15 (1.81)	3.41 (1.97)	–0.769[a]	.44
Physical function, median (IQR)	0.60 (1.70)	0.60 (1.90)	0.60 (1.50)	–0.586[c]	.56
Adherence to exercise, per week , n (%)	—	—	—	10.450[b]	.06
0	9 (7.6)	1 (1.7)	8 (13.6)	—	—
Occasionally	66 (55.9)	34 (57.6)	32 (54.2)	—	—

Characteristic	Total (n=118)	Intervention (n=59)	Control (n=59)	$t^a/\chi^{2b}/Z^c$	*P* value
1 or 2 days	14 (11.9)	9 (15.3)	5 (8.5)	—	—
3 or 4 days	11 (9.3)	8 (13.6)	3 (5.1)	—	—
5 or 6 days	1 (0.8)	0 (0)	1 (1.7)	—	—
Daily	17 (14.4)	7 (11.9)	10 (16.9)	—	—

[a]independent sample *t* test.

[b]chi-square test.

[c]Mann-Whitney *U* test.

[d]Not applicable.

[e]AS: ankylosing spondylitis.

Primary Outcomes

The correct item score, correct answer score, and 4 area scores of knowledge of AS and self-efficacy scores in the intervention group were significantly higher than the control group after the intervention (all *P*<.001). A larger proportion of participants in the intervention group adhered to regular exercise after the intervention compared with the control group (*P*=.003, Table 3). The within-group analyses for the intervention group showed significant increases in all scores of AS knowledge and self-efficacy scores (all *P*<.001). The within-group analyses for the control group detected increases in correct answer score (*P*=.04) and general knowledge score (*P*=.002), but no significant difference in self-efficacy score, other scores of knowledge of AS including the correct item score (all *P*>.05, Table 4).

Table 3. Comparison of outcomes between groups at posttest.

Characteristic	Intervention (n=59)	Control (n=59)	$t^a/Z^b/\chi^{2c}$	*P* value
Knowledge of AS[d]				
Correct item score, mean (SD)	11.81 (2.44)	6.83 (3.34)	9.249[a]	<.001[e]
Correct answer score, mean (SD)	22.49 (3.35)	16.05 (5.17)	8.022[a]	<.001[e]
General knowledge, mean (SD)	7.20 (1.31)	5.80 (1.75)	4.943[a]	<.001[e]
B27 antigen and inheritance, median (IQR)	2.00 (1.00)	1.00 (1.00)	–7.139[b]	<.001[e]
Drug treatment and physical therapy, mean (SD)	8.48 (1.01)	5.95 (2.31)	7.706[a]	<.001[e]
Joint protection, pacing and priorities, mean (SD)	4.53 (0.80)	3.39 (1.29)	5.766[a]	<.001[e]
Self-efficacy, mean (SD)	7.60 (1.50)	6.41 (2.04)	3.612[a]	<.001[e]
Disease activity, mean (SD)	2.95 (1.71)	3.41 (1.76)	–1.434[a]	.15
Physical function, median (IQR)	1.00 (1.40)	1.40 (1.60)	–1.764[b]	.08
Adherence to exercise, per week, n (%)	—[f]	—	18.028[c]	.003[e]
0	1 (1.7)	6 (10.2)	—	—
Occasionally	11 (18.6)	28 (47.5)	—	—
1 or 2 days	15 (25.4)	8 (13.6)	—	—
3 or 4 days	12 (20.3)	7 (11.9)	—	—
5 or 6 days	6 (10.2)	4 (6.8)	—	—
Daily	14 (23.7)	6 (10.2)	—	—

[a]independent sample *t* test.

[b]Mann-Whitney *U* test.

[c]chi-square test.

[d]AS: ankylosing spondylitis.

[e]*P*<.01.

[f]Not applicable.

Table 4. Comparison of outcomes within the intervention and control groups.

Characteristic and group	Pretest, mean (SD)	Posttest, mean (SD)	Difference of means (95% CI)	t [a]	P value
Knowledge of AS[b]					
Correct item score					
IG[c]	6.64 (2.92)	11.81 (2.44)	5.17 (4.22 to 6.11)	10.953	<.001[d]
CG[e]	6.17 (3.24)	6.83 (3.34)	0.66 (–0.12 to 1.44)	1.694	.10
Correct answer score					
IG	15.61 (4.58)	22.49 (3.35)	6.88 (5.55 to 8.21)	10.345	<.001[d]
CG	14.97 (5.24)	16.05 (5.17)	1.08 (0.06 to 2.11)	2.126	.04[f]
General knowledge					
IG	5.19 (1.82)	7.20 (1.31)	2.02 (1.52 to 2.51)	8.123	<.001[d]
CG	5.23 (1.91)	5.80 (1.75)	0.56 (0.21 to 0.91)	3.231	.002[d]
B27 antigen and inheritance					
IG	0.75 (0.76)	2.29 (0.85)	1.54 (1.28 to 1.80)	11.988	<.001[d]
CG	0.86 (0.75)	0.92 (0.75)	0.05 (–0.17 to 0.27)	0.465	.64
Drug treatment and physical therapy					
IG	6.34 (1.91)	8.47 (1.01)	2.14 (1.59 to 2.68)	7.794	<.001[d]
CG	5.76 (2.37)	5.95 (2.31)	0.19 (–0.22 to 0.59)	0.919	.36
Joint protection, pacing, and priorities					
IG	3.34 (1.17)	4.53 (0.80)	1.19 (0.84 to 1.53)	6.840	<.001[d]
CG	3.10 (1.52)	3.39 (1.29)	0.29 (–0.12 to 0.70)	1.404	.17
Self-efficacy					
IG	6.41 (1.91)	7.60 (1.50)	1.19 (0.72 to 1.66)	5.055	<.001[d]
CG	6.18 (1.97)	6.41 (2.04)	0.22 (–0.16 to 0.62)	1.178	.24
Disease activity					
IG	3.15 (1.81)	2.96 (1.71)	–0.19 (–0.70 to 0.32)	–0.754	.45
CG	3.41 (1.97)	3.41 (1.76)	0 (–0.52 to 0.52)	–0.005	>.99
Physical function					
IG	1.61 (2.23)	1.67 (1.79)	0.07 (–0.44 to 0.57)	0.269	.79
CG	1.25 (1.63)	1.91 (1.65)	0.67 (0.26 to 1.07)	3.320	.002[d]

[a]paired sample t test.

[b]AS: ankylosing spondylitis.

[c]IG: intervention group.

[d]P<.01.

[e]CG: control group.

[f]P<.05.

Secondary Outcomes

At posttest, there was no difference in disease activity or physical function between the intervention and control groups (P>.05). The within-group analyses for the intervention group showed no significant change in disease activity or physical function (P>.05). The within-group analyses for the control group detected a decline in physical function (P=.002), but no significant change in disease activity (P>.05, Table 3 and Table 4).

Discussion

Principal Findings

This study explored the effects of the theory-based educational intervention through WeChat among Chinese patients with AS.

Our findings demonstrated that this intervention was feasible and beneficial for improving patient disease knowledge, self-efficacy, and exercise adherence, which was in line with previous studies [12,16,41]. The educational intervention delivered by WeChat can increase access to health care providers for participants, teach knowledge and skill of disease management, and have positive effects in AS patients.

We found that the theory-based educational intervention can increase patient knowledge about AS, which was in line with prior studies [16,41], and corresponds with educating the patients on perceived severity of the disease and perceived susceptibility to increased limited mobility without action. Haglund et al [30] revealed that 43% of spondyloarthritis patients had educational needs. Moreover, patient knowledge levels of AS in this study were relatively low compared with previous studies [36]. In our study, the research nurses provided knowledge of managing AS, which may increase patient knowledge levels of AS.

The educational intervention via WeChat can effectively improve self-efficacy of AS patients, a finding similar to prior studies [18,45]. Self-efficacy can be enhanced through direct experience, alternative experience, and verbal persuasion [46]. In this study, participants gained knowledge and peer experience about disease management through educational information and nurse coaching by praising small accomplishments, which increased their perceived self-efficacy. Learning about useful experiences of others can inspire patients to try strategies to manage disease [47]. Seeing the adaptations of others to AS helps patients manage their disease better and improves confidence in coping with disease, which increases their self-efficacy to manage the disease. Persuasion from research nurses can help patients successfully manage their conditions [48]. That these skills may enhance patient confidence in managing disease and improving their self-efficacy was shown in our intervention.

This intervention effectively improved patient self-efficacy which, in turn, may have contributed to higher adherence to exercise. In this study, a larger proportion of participants in the intervention group adhered to regular exercise compared with the control group after the intervention. The finding was in line with earlier studies [16,41]. Self-efficacy is an important factor influencing exercise behavior in AS patients [49]. Our intervention helped patients perceive the severe consequences of AS, educated them on the importance of disease management, and taught them skills to manage their condition, which may have prompted regular exercise and helped them develop cues to action in their daily lives. In addition, the intervention delivered through WeChat may make it easier to exercise at home. These issues may enhance patient exercise adherence.

The results of this study did not detect significant differences in disease activity and physical function except for a decline in physical function in the control group. Previous reports on the efficacy of patient education on disease activity and physical function are inconsistent [14-18,20]. In our study, patients had relatively low disease activity and functional limitation, and these variables may be difficult to modify. Our 12-week intervention period may not be long enough to detect significant changes in biomarkers, such as disease activity, function, etc. Educational intervention may not produce a direct effect on health outcomes [2]. Thus, future studies should explore the long-term effects of educational intervention on health outcomes.

Limitations

This study had several strengths. An assessor blinded to group assignment collected pretest and posttest data to reduce biased responses. Furthermore, using HBM might increase the efficacy of this intervention. Finally, we used an intention-to-treat analysis with multiple imputations for missing data to reduce bias in assessment of treatment effects.

This study had several potential limitations. First, we only recruited patients from a tertiary hospital who were able to use WeChat. Although the use of smartphone and internet access are relatively ubiquitous, the use of WeChat limits the generalization of findings to all Chinese AS patients. Second, patient views and cost-effectiveness analysis are important to evaluate and improve this educational intervention, but we did not collect these data because of limited time and financial support. Third, we collected outcomes at 12 weeks, but the effects of WeChat-based education on health outcomes may only become apparent in a long-term.

Conclusions

We demonstrated that the theory-based educational intervention delivered through WeChat, led by experienced nurses, was feasible and effective to improve AS patient disease knowledge, self-efficacy, and exercise adherence in a short-term. WeChat can deliver timely health service for patients with no available time or living in rural communities. During the COVID-19 pandemic period, the intervention approach may help health care providers provide continuous rheumatology care. We suggest that this intervention can be integrated into routine rheumatology care. Future studies should explore long-term effects of this intervention.

Acknowledgments

We would like to thank all patients who participated in this study.

Authors' Contributions

YS and HC were responsible for conceptualization. YS, YC, YW, and HC were responsible for methodology. YS and YC provided validation. YS performed data curation. YS and ER provided formal analysis and wrote the original draft. YC and YW were responsible for investigation. HC supervised the project. All authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 13514 KB - jmir_v24i10e38501_app1.pdf]

References

1. van der Heijde D, Ramiro S, Landewé R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017 Jun 13;76(6):978-991. [doi: 10.1136/annrheumdis-2016-210770] [Medline: 28087505]
2. Zangi HA, Ndosi M, Adams J, Andersen L, Bode C, Boström C, European League Against Rheumatism (EULAR). EULAR recommendations for patient education for people with inflammatory arthritis. Ann Rheum Dis 2015 Jun 03;74(6):954-962 [FREE Full text] [doi: 10.1136/annrheumdis-2014-206807] [Medline: 25735643]
3. Braun J. Axial spondyloarthritis including ankylosing spondylitis. Rheumatology (Oxford) 2018 Aug 01;57(suppl_6):vi1-vi3 [FREE Full text] [doi: 10.1093/rheumatology/key079] [Medline: 30445482]
4. Wang Y, Liu X, Wang W, Shi Y, Ji X, Hu L, et al. Adherence, efficacy, and safety of wearable technology-assisted combined home-based exercise in chinese patients with ankylosing spondylitis: randomized pilot controlled clinical trial. J Med Internet Res 2022 Jan 18;24(1):e29703 [FREE Full text] [doi: 10.2196/29703] [Medline: 35040798]
5. McDonald MT, Siebert S, Coulter EH, McDonald DA, Paul L. Level of adherence to prescribed exercise in spondyloarthritis and factors affecting this adherence: a systematic review. Rheumatol Int 2019 Feb 16;39(2):187-201 [FREE Full text] [doi: 10.1007/s00296-018-4225-8] [Medline: 30649569]
6. Smolen JS, Gladman D, McNeil HP, Mease PJ, Sieper J, Hojnik M, et al. Predicting adherence to therapy in rheumatoid arthritis, psoriatic arthritis or ankylosing spondylitis: a large cross-sectional study. RMD Open 2019 Jan 11;5(1):e000585 [FREE Full text] [doi: 10.1136/rmdopen-2017-000585] [Medline: 30713716]
7. Raghunath S, Hijjawi R, Hoon E, Shanahan EM, Goldblatt F. Qualitative assessment of medication adherence in patients with rheumatic diseases on biologic therapy. Clin Rheumatol 2019 Oct 6;38(10):2699-2707. [doi: 10.1007/s10067-019-04609-y] [Medline: 31172365]
8. Sang Y, Dong C, Fu T, Zhao R, Ge X, Zhou W, et al. Associated factors with adherence to standard exercise therapy and health-related quality of life in Chinese patients with ankylosing spondylitis. Mod Rheumatol 2020 Jan 04;30(1):149-154 [FREE Full text] [doi: 10.1080/14397595.2018.1559966] [Medline: 30561239]
9. Schrieber L, Colley M. Patient education. Best Pract Res Clin Rheumatol 2004 Aug;18(4):465-476. [doi: 10.1016/j.berh.2004.03.008] [Medline: 15301981]
10. Ndosi M, Johnson D, Young T, Hardware B, Hill J, Hale C, et al. Effects of needs-based patient education on self-efficacy and health outcomes in people with rheumatoid arthritis: a multicentre, single blind, randomised controlled trial. Ann Rheum Dis 2016 Jun 10;75(6):1126-1132 [FREE Full text] [doi: 10.1136/annrheumdis-2014-207171] [Medline: 26162769]
11. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
12. Uritani D, Koda H, Sugita S. Effects of self-management education programmes on self-efficacy for osteoarthritis of the knee: a systematic review of randomised controlled trials. BMC Musculoskelet Disord 2021 Jun 05;22(1):515 [FREE Full text] [doi: 10.1186/s12891-021-04399-y] [Medline: 34090406]
13. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 01;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
14. Kaya T, Goksel Karatepe A, Atici Ozturk P, Gunaydin R. Impact of peer-led group education on the quality of life in patients with ankylosing spondylitis. Int J Rheum Dis 2016 Feb;19(2):184-191. [doi: 10.1111/1756-185X.12256] [Medline: 24330320]
15. Kasapoglu Aksoy M, Birtane M, Taştekin N, Ekuklu G. The effectiveness of structured group education on ankylosing spondylitis patients. J Clin Rheumatol 2017 Apr;23(3):138-143. [doi: 10.1097/RHU.0000000000000499] [Medline: 28248799]
16. Rodríguez-Lozano C, Juanola X, Cruz-Martínez J, Peña-Arrébola A, Mulero J, Gratacós J, Spondyloarthropathies Study Group of the Spanish Society of Rheumatology. Outcome of an education and home-based exercise programme for patients with ankylosing spondylitis: a nationwide randomized study. Clin Exp Rheumatol 2013;31(5):739-748. [Medline: 23899791]
17. O'Dwyer T, Monaghan A, Moran J, O'Shea F, Wilson F. Behaviour change intervention increases physical activity, spinal mobility and quality of life in adults with ankylosing spondylitis: a randomised trial. J Physiother 2017 Jan;63(1):30-39 [FREE Full text] [doi: 10.1016/j.jphys.2016.11.009] [Medline: 27989730]

18. Inderjeeth CA, Boland E, Connor C, Johnson C, Jacques A, McQuade J. Evaluation of an ankylosing spondylitis education and self-management program: beneficial effects on ankylosing spondylitis specific outcomes. Int J Rheum Dis 2021 Mar;24(3):434-444. [doi: 10.1111/1756-185X.14070] [Medline: 33694321]
19. Molto A, Gossec L, Poiraudeau S, Claudepierre P, Soubrier M, Fayet F, et al. Evaluation of the impact of a nurse-led program of systematic screening of comorbidities in patients with axial spondyloarthritis: the results of the COMEDSPA prospective, controlled, one year randomized trial. Semin Arthritis Rheum 2020 Aug;50(4):701-708. [doi: 10.1016/j.semarthrit.2020.05.012] [Medline: 32521324]
20. Molto A, Gossec L, Poiraudeau S, Claudepierre P, Soubrier M, Fayet F, et al. Evaluation of the impact of a nurse-led program of patient self-assessment and self-management in axial spondyloarthritis: results of a prospective, multicentre, randomized, controlled trial (COMEDSPA). Rheumatology (Oxford) 2021 Feb 01;60(2):888-895. [doi: 10.1093/rheumatology/keaa480] [Medline: 33063096]
21. Ji X, Hu L, Wang Y, Luo Y, Zhu J, Zhang J, et al. "Mobile health" for the management of spondyloarthritis and its application in China. Curr Rheumatol Rep 2019 Nov 19;21(11):61. [doi: 10.1007/s11926-019-0860-7] [Medline: 31741083]
22. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
23. Li X, Li T, Chen J, Xie Y, An X, Lv Y, et al. A WeChat-based self-management intervention for community middle-aged and elderly adults with hypertension in Guangzhou, China: a cluster-randomized controlled trial. Int J Environ Res Public Health 2019 Oct 23;16(21):4058 [FREE Full text] [doi: 10.3390/ijerph16214058] [Medline: 31652688]
24. Jao N. WeChat now has over 1 billion active monthly users worldwide. TechNode. 2018 Mar 05. URL: https://technode.com/2018/03/05/wechat?1?billion?users/ [accessed 2022-09-28]
25. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):e117. [doi: 10.2196/jmir.8382] [Medline: 29592845]
26. Zhou K, Wang W, Zhao W, Li L, Zhang M, Guo P, et al. Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: a clinical randomized controlled trial. Int J Nurs Stud 2020 Jun;106:103565. [doi: 10.1016/j.ijnurstu.2020.103565] [Medline: 32272281]
27. Ren Q, Lian M, Liu Y, Thomas-Hawkins C, Zhu L, Shen Q. Effects of a transtheoretical model-based WeChat health education programme on self-management among haemodialysis patients: a longitudinal experimental intervention study. J Adv Nurs 2019 Dec 13;75(12):3554-3565. [doi: 10.1111/jan.14182] [Medline: 31441525]
28. Ni Z, Wu B, Yang Q, Yan LL, Liu C, Shaw RJ. An mHealth intervention to improve medication adherence and health outcomes among patients with coronary heart disease: randomized controlled trial. J Med Internet Res 2022 Mar 09;24(3):e27202 [FREE Full text] [doi: 10.2196/27202] [Medline: 35262490]
29. Ni Z, Liu C, Wu B, Yang Q, Douglas C, Shaw RJ. An mHealth intervention to improve medication adherence among patients with coronary heart disease in China: development of an intervention. Int J Nurs Sci 2018 Oct 10;5(4):322-330 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.003] [Medline: 31406843]
30. Haglund E, Bremander A, Bergman S, Larsson I. Educational needs in patients with spondyloarthritis in Sweden: a mixed-methods study. BMC Musculoskelet Disord 2017 Aug 02;18(1):335 [FREE Full text] [doi: 10.1186/s12891-017-1689-8] [Medline: 28768510]
31. Rosenstock IM. Historical origins of the health belief model. Health Educ Monog 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
32. Sremanakova J, Sowerbutts AM, Todd C, Cooke R, Burden S. Systematic review of behaviour change theories implementation in dietary interventions for people who have survived cancer. Nutrients 2021 Feb 13;13(2):612 [FREE Full text] [doi: 10.3390/nu13020612] [Medline: 33668596]
33. Parwati NM, Bakta IM, Januraga PP, Wirawan IMA. A health belief model-based motivational interviewing for medication adherence and treatment success in pulmonary tuberculosis patients. Int J Environ Res Public Health 2021 Dec 15;18(24):13238 [FREE Full text] [doi: 10.3390/ijerph182413238] [Medline: 34948846]
34. Song Y, Xie X, Chen Y, Wang Y, Yang H, Nie A, et al. The effects of WeChat-based educational intervention in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Res Ther 2021 Mar 04;23(1):72 [FREE Full text] [doi: 10.1186/s13075-021-02453-7] [Medline: 33663558]
35. Orji R, Vassileva J, Mandryk R. Towards an effective health interventions design: an extension of the health belief model. Online J Public Health Inform 2012 Dec 19;4(3):ojphi.v4i3.4321 [FREE Full text] [doi: 10.5210/ojphi.v4i3.4321] [Medline: 23569653]
36. Claudepierre P, Flipo R, Sibilia J, Berthelot J, Goupille P, Cortinovis S, et al. Patient knowledge of their disease: a French multicenter study in ankylosing spondylitis. Joint Bone Spine 2004 Nov;71(6):550-556. [doi: 10.1016/j.jbspin.2004.09.001] [Medline: 15589438]
37. Lubrano E, Helliwell P, Moreno P, Griffiths B, Emery P, Veale D. The assessment of knowledge in ankylosing spondylitis patients by a self-administered questionnaire. Br J Rheumatol 1998 Apr;37(4):437-441. [doi: 10.1093/rheumatology/37.4.437] [Medline: 9619896]

38. Gao L, Zhang X, Li M, Yuan J, Cui X, Shi B. Psychometric properties of the Chinese version of Arthritis Self-Efficacy Scale-8 (ASES-8) in a rheumatoid arthritis population. Rheumatol Int 2017 May 6;37(5):751-756. [doi: 10.1007/s00296-016-3640-y] [Medline: 28063069]
39. Wilcox S, Schoffman DE, Dowda M, Sharpe PA. Psychometric properties of the 8-item english arthritis self-efficacy scale in a diverse sample. Arthritis 2014 Aug 21;2014:385256 [FREE Full text] [doi: 10.1155/2014/385256] [Medline: 25215233]
40. Sandhu J, Packham JC, Healey EL, Jordan KP, Garratt AM, Haywood KL. Evaluation of a modified arthritis self-efficacy scale for an ankylosing spondylitis UK population. Clin Exp Rheumatol 2011;29(2):223-230. [Medline: 21504660]
41. Sudre A, Figuereido IT, Lukas C, Combe B, Morel J. On the impact of a dedicated educational program for ankylosing spondylitis: effect on patient satisfaction, disease knowledge and spinal mobility, a pilot study. Joint Bone Spine 2012 Jan;79(1):99-100. [doi: 10.1016/j.jbspin.2011.06.005] [Medline: 21852173]
42. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994 Dec 21;21(12):2286-2291. [Medline: 7699630]
43. Calin A, Garrett S, Whitelock H, Kennedy L, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994 Dec;21(12):2281-2285. [Medline: 7699629]
44. wjx.cn. URL: https://www.wjx.cn/ [accessed 2022-10-07]
45. Zhao S, Chen H. Effectiveness of health education by telephone follow-up on self-efficacy among discharged patients with rheumatoid arthritis: a randomised control trial. J Clin Nurs 2019 Nov 28;28(21-22):3840-3847. [doi: 10.1111/jocn.15002] [Medline: 31325348]
46. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191]
47. Grønning K, Skomsvoll JF, Rannestad T, Steinsbekk A. The effect of an educational programme consisting of group and individual arthritis education for patients with polyarthritis: a randomised controlled trial. Patient Educ Couns 2012 Jul;88(1):113-120. [doi: 10.1016/j.pec.2011.12.011] [Medline: 22277625]
48. Jiao N, Zhu L, Chong YS, Chan WS, Luo N, Wang W, et al. Web-based versus home-based postnatal psychoeducational interventions for first-time mothers: a randomised controlled trial. Int J Nurs Stud 2019 Nov;99:103385. [doi: 10.1016/j.ijnurstu.2019.07.002] [Medline: 31442783]
49. Lim H, Lim H, Lee MS. Relationship between self-efficacy and exercise duration in patients with ankylosing spondylitis. Clin Rheumatol 2005 Aug 24;24(4):442-443. [doi: 10.1007/s10067-004-0974-8] [Medline: 15338448]

Abbreviations

BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
AS: ankylosing spondylitis
ASES-8: Arthritis Self-Efficacy Scale–8
HBM: health belief model
mHealth: mobile health

Edited by R Kukafka; submitted 05.04.22; peer-reviewed by Z Ni, UH Mohamad; comments to author 26.04.22; revised version received 14.05.22; accepted 26.09.22; published 20.10.22.

Please cite as:
Song Y, Reifsnider E, Chen Y, Wang Y, Chen H
The Impact of a Theory-Based mHealth Intervention on Disease Knowledge, Self-efficacy, and Exercise Adherence Among Ankylosing Spondylitis Patients: Randomized Controlled Trial
J Med Internet Res 2022;24(10):e38501
URL: https://www.jmir.org/2022/10/e38501
doi:10.2196/38501
PMID:36264605

Original Paper

The Efficacy of a Web-Based Stress Management Intervention for Employees Experiencing Adverse Working Conditions and Occupational Self-efficacy as a Mediator: Randomized Controlled Trial

Patricia Nixon[1], MSc; David Daniel Ebert[2], PhD; Leif Boß[1], PhD; Peter Angerer[3], PhD; Nico Dragano[4], PhD; Dirk Lehr[1], PhD

[1]Department of Health Psychology and Applied Biological Psychology, Institute of Psychology, Leuphana University, Lueneburg, Germany

[2]Division of Psychology & Digital Mental Health Care, Department for Sport & Health Sciences, Technical University of Munich, Munich, Germany

[3]Faculty of Medicine, Institute of Occupational, Social and Environmental Medicine, Centre for Health and Society, Heinrich Heine University, Duesseldorf, Germany

[4]Institute of Medical Sociology, Centre for Health and Society, University Hospital, Duesseldorf, Germany

Corresponding Author:
Dirk Lehr, PhD
Department of Health Psychology and Applied Biological Psychology
Institute of Psychology
Leuphana University
Universitätsallee 1
Lueneburg, 21335
Germany
Phone: 49 41316772720
Email: lehr@leuphana.de

Abstract

Background: Work stress is highly prevalent and puts employees at risk for adverse health consequences. Web-based stress management interventions (SMIs) promoting occupational self-efficacy might be a feasible approach to aid employees to alleviate this burden and to enable them to improve an unbalanced situation between efforts and rewards at work.

Objective: The first aim of this randomized controlled trial was to investigate the efficacy of a web-based SMI for employees perceiving elevated stress levels and an effort-reward imbalance in comparison to a waitlist control (WLC) group. Second, we investigated whether the efficacy of an SMI could be explained by an increase in occupational self-efficacy and whether this personal resource enables employees to change adverse working conditions.

Methods: A total of 262 employees reporting effort-reward imbalance scores over 0.715 and elevated stress levels (10-item Perceived Stress Scale [PSS-10] score ≥22) were randomly assigned to either the intervention group (IG; SMI) or the WLC group. The primary outcome was perceived stress measured using the PSS-10. The secondary outcomes included mental and work-related health measures. Four different mediation analyses were conducted with occupational self-efficacy, efforts, and rewards as mediators. After eligibility screening, data were collected web based at baseline (T1), 7 weeks (T2) and 6 months (T3).

Results: Study participation was completed by 80% (105/130, 80.8%) in the IG and 90% (119/132, 90.2%) in the WLC group. Analyses of covariance revealed that stress reduction was significantly higher for the SMI group compared with the WLC group at T2 (d=0.87, 95% CI 0.61-1.12, P<.001) and T3 (*d*=0.65, 95% CI 0.41-0.90, P<.001). Mediation analyses indicated that occupational self-efficacy mediated the beneficial effect of the SMI on stress directly. Furthermore, the analyses revealed a significant indirect effect of occupational self-efficacy via rewards (*b*=0.18, t_{259}=4.52, P<.001), but not via efforts (b=0.01, t_{259}=0.27, P>.05) while efforts still had a negative impact on stress (*b*=0.46, t_{257}=2.32, P<.05).

Conclusions: The SMI was effective in reducing stress and improving occupational self-efficacy in employees despite them experiencing an effort-reward imbalance at work. Results from mediation analyses suggest that fostering personal resources such as occupational self-efficacy contributes to the efficacy of the SMI and enables employees to achieve positive changes regarding the rewarding aspects of the workplace. However, the SMI seemed to neither directly nor indirectly impact efforts, suggesting

that person-focused interventions might not be sufficient and need to be complemented by organizational-focused interventions to comprehensively improve mental health in employees facing adverse working conditions.

Trial Registration: German Clinical Trials Register DRKS00005990; https://tinyurl.com/23fmzfu3

(J Med Internet Res 2022;24(10):e40488) doi:10.2196/40488

KEYWORDS

occupational eMental health; stress; occupational self-efficacy; effort-reward imbalance; randomized controlled trial

Introduction

More than a decade ago, the World Health Organization identified stress as major risk factor for adverse consequences on physical and mental health for the 21st century [1]. In particular, the workplace can be a source of stress that can be associated with an increased risk of depression and cardiovascular diseases [2-4]. Next to such harmful effects on employees' personal lives and health, experiencing high strain at work can entail substantial societal costs [5].

One of the most prominent theoretical frameworks to investigate workplace stressors is the effort-reward imbalance model [6]. In short, this model is based on the premise of an imbalance between efforts invested and low rewards received in return. Both efforts and rewards therefore reflect subjectively perceived working conditions employees are exposed to. Rewards can be distinguished between financial payments, job security, or career prospects, and intangible compensation such as esteem or praise. During the past few decades, the model was well researched. It was shown that employees experiencing an effort-reward imbalance have an increased risk of depression [7], lower immunity [8], or coronary heart disease [9]. Multiple systematic reviews demonstrated robust evidence for the links between an effort-reward imbalance and health, and suggested that it can instigate psychological, physical, and behavioral health–impairing pathways [10-12].

Psychosocial hazards were identified as one of the key emerging health risks [13] and there were significant developments to address psychosocial risk factors at work. For example, the National Standard of Canada for Psychological Health and Safety in the Workplaces provides a comprehensive framework for an approach to ensure a psychologically healthy workplace [14]. Notably, changing adverse working conditions requires a timely and complex transformational process that can be a considerable source of work stress itself and is associated with different risks such as an increase of stress-related medication intake [15].

Workers already affected by high levels of stress are in an acute need of relief, and so waiting for the successful implementation of organizational changes can be challenging. In this situation, a stress management intervention (SMI) might be a first step to support those in need of help sooner [16]. There is evidence for the beneficial effects of SMIs in traditional face-to-face settings [17], which was complemented by a more recent and growing body of research for the web-based delivery [16,18,19]. Web-based interventions allow the workforce to benefit from low-threshold access and highly flexible participation in terms of time and location, and employers to profit from easy scalability and low required resources [20]. Furthermore, they might have the potential to alleviate the burden of workplace stressors by promoting self-efficacy and improving various health outcomes such as insomnia or depression [21-24] in both short and long term [25,26]. However, until today, evidence is missing on whether a web-based SMI could also effectively reduce perceived stress in employees who are exposed to adverse working conditions in terms of an imbalance between efforts and rewards. Moreover, no trial has yet examined mechanisms of change within this high-risk population and whether an increase in personal resources could enable employees to improve the unbalanced situation between efforts and rewards at work.

An effective implementation of a web-based SMI for employees who are exposed to adverse working conditions could be a person-centered intervention helping workers to initiate changes, a strategy known as problem-focused coping following the transactional model by Lazarus and Folkman [27]. A necessary personal resource for self-initiated changes employees make to redesign working conditions is self-efficacy, which is believed to trigger proactive behaviors undertaken at work [28]. Initially, Bandura [29] defined self-efficacy as confidence to meet difficult challenges or prospective problems by oneself. Individuals with high self-efficacy experience lower levels of work strain and engage more in problem-focused coping [30]. Another study confirmed that a problem-solving training for teachers could strengthen the ability to cope with problems and stressful situations as well as increase self-efficacy [31]. Within this organizational context, occupational self-efficacy can be described as personal belief in work-related abilities [32]. Studies on occupational self-efficacy have demonstrated positive associations with job performance, employee satisfaction, employability, and work commitment, and negative relationships with job insecurity [32,33]. A study on the same SMI that was examined in this randomized controlled trial (RCT) provided first evidence for effects on occupational self-efficacy [34], while the previously stated need for research on self-efficacy as a mechanism of change in an occupational SMI has not been addressed yet [35]. Moreover, there is no evidence on the effects of occupational self-efficacy on the perception of adverse working conditions yet despite the assumption that self-efficacy as a function of self-regulation conducive to health relies on successful exchange of efforts and rewards [36].

To the best of our knowledge, this is the first RCT to investigate the efficacy of a web-based occupational SMI in employees perceiving high stress levels and an effort-reward imbalance and to explore mediating effects of occupational self-efficacy, efforts, and rewards on stress reduction. This trial will examine the hypothesis that the SMI will effectively reduce perceived

stress in the intervention group (IG) compared with a waitlist control (WLC) group. The second study aim is to investigate mediating effects of the personal resource of occupational self-efficacy and environmental factors, specifically efforts, and rewards at the workplace in the association between the intervention and perceived stress.

Methods

Study Design and Conditions

A primary RCT including 264 participants experiencing an effort-reward imbalance was conducted in compliance with the Declaration of Helsinki and Good Clinical Practice and following the CONSORT (Consolidated Standards of Reporting Trials) guidelines [37,38] (Multimedia Appendix 1). Based on meta-analytic evidence for web-based SMI revealing moderate effects (Hedges g=0.54) [19] and considering the impact of adverse working conditions, this study aimed to detect differences between groups with an effect size of Cohen d=0.35 based on a power (1–β) of 0.80 in a 2-tailed test with α=.05. Participants were randomly assigned to the IG or the WLC group at a ratio of 1:1 using an automated computer-based random integer generator (DatInf RandList; Datinf GmbH). Participants were allocated to the study groups by an independent researcher not otherwise involved in the study. Self-reported outcomes were assessed between May 2014 and May 2015 with a secured online-based self-report system (AES; 256-bit encrypted) at screening for eligibility (T0), baseline (T1), and 7 weeks (T2), and 6 months (T3) after randomization. After allocation, participants in the IG received immediate access to the intervention, whereas those in the WLC group obtained access after 6 months. Treatment as usual was not restricted and monitored. None of the obtained data presented here were published before.

Participants and Recruitment

Participants were recruited from the general working population via the research project website and mass media (eg, articles in health insurance magazines). Inclusion criteria were the willingness to give informed consent; legal age (18 years); employment; 10-item Perceived Stress Scale (PSS-10) [39,40] score ≥22; effort-reward imbalance [41] score >0.715, which was found to indicate a highly hazardous imbalance between effort and rewards at the workplace [42]; no notable suicidal risk, as indicated by a score of >1 on item 9 (I feel I would be better off dead) of the Beck Depression Inventory [43]; and no previous or current diagnosis of dissociative symptoms or psychosis. Interested participants signed up on the open access website with their email address to receive a link to the eligibility screening questionnaire. Eligible applicants were required to provide informed consent and baseline data (T1).

Intervention

Psychologists developed the intervention for employees based on Lazarus' transactional model of stress focusing on problem solving and emotion regulation skills [27]. The intervention encouraged participants to reflect on meaningful issues that were not restricted to either work or personal life. The efficacy was demonstrated before in an indicated prevention sample and with different guidance formats, namely, adherence-focused guidance and self-help [22,44,45]. The SMI consisted of 7 core modules and an optional booster session 4 weeks after termination. Module completion required 45-60 minutes and participants were advised to complete at least one per week, adding up to an intervention period of 4-7 weeks. Participants could choose whether and how often they preferred to receive short automatic motivational SMS text messages to their mobile device (infrequent or intensive, ie, 1-3 SMS text messages daily). In addition, participants could inquire feedback-on-demand, which was provided by an e-coach within 48 hours only upon request on the internal messaging platform. E-coaches were skilled psychologists following feedback guidelines from the standardized manual for the intervention. Participants were assigned to an e-coach in a 1-to-1 ratio.

Primary Outcome Measure

The primary outcome was perceived stress appraised with the German version of PSS-10 [39,40], which was also developed based on Lazarus' transactional model of stress. The items assess to what extent participants experienced their lives as stressful within the past week on a 5-point Likert scale from 0 (never) to 4 (very often), resulting in sums from 0 to 40, with higher scores reflecting higher levels of stress. In this study, values for the internal reliability (Cronbach α) were .81 at T1, .89 at T2, and .92 at T3.

Secondary Outcome Measures

Included measures for the secondary outcomes are listed in the following sections, with number of items, item range, and reliabilities assessed at T2.

Mediators

Among the secondary outcomes, 2 measures were assessed for the inclusion as mediators. First, the Effort Reward Imbalance Questionnaire Short Form [41] with the subscales *efforts* (3 items; α=.78) and *rewards* (7 items; α=.79; score range 1-4). And second, the short form of the Occupational Self-Efficacy Scale (OSS-SF [32]; 6 items; α=.89; score range 1-6).

Work-Related Health

The subscale *emotional exhaustion* of the Maslach Burnout Inventory (MBI-GS-D; 5 items; α=.87; score range 1-6) was used to evaluate work-related health [46]. The Utrecht Working Scale (UWES) [47] was used to examine work engagement (9 items; α=.93; score range 0-6). A single-item question was used to assess work ability (Work Ability Index) [48] and the Work Limitations Questionnaire [49] was administered to examine presenteeism.

Mental Health

The short version of the Centre for Epidemiological Studies' Depression scale (CES-D) [50,51] was used to assess depression (15 items; α=.84; score range 0-3). The Connor-Davidson Resilience Scale [52] was used to examine resilience (10 items; α=.88; score range 0-4). The Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument [53] was used to examine health-related quality of life (35 items, different ranges from 1 to 5 and 1 to 6; α=.96) at T3.

Other Measures

To assess the level of satisfaction with the intervention, the Client Satisfaction Questionnaire adapted to web-based interventions was used (CSQ-I; 8 items; α=.92; score range 1-4) [54]. In addition, self-developed measures were used to assess demographics, current occupation, work sector, income, educational level, and previous use of health services.

Statistical Analyses

Statistical analyses were performed according to the recommendations of the CONSORT statement [37]. Data were analyzed with SPSS Statistics version 25 (IBM Corp.) [55] based on the intention-to-treat principle. An additional per-protocol analysis was conducted for the primary outcome, including only participants who completed at least six modules. Analyses of covariance (ANCOVA) were calculated with outcome baseline scores as covariates and a 2-tailed significance level at P<.05 to detect between-group differences for the IG and the WLC group at T2 and T3. Simulation studies have already demonstrated the methodological robustness of ANCOVA against bias, higher precision, and statistical power for experimental studies [56,57]. To handle missing data, multiple imputations were conducted for the intention-to-treat and per-protocol analyses with 10 estimates for each value that were aggregated into an overall value [58].

Response Analyses

The Reliable Change Index of Jacobson and Truax [59] was used to investigate improvements of the primary outcome on an individual level. The SD of 6.2 and the reliability of PSS-10 of the norm population [60] were used in the formula [1.96 × SD1 × sqrt(2) × sqrt(1–rel)] to calculate that a reduction in perceived stress could be defined as *reliably improved* if changes of more than ±5.16 points were detected from T1 to T2. Symptom-free status was achieved according to Jacobson and Truax [59] when participants scored more than 2 SDs below the baseline mean (T1) of the primary outcome in the IG (mean 23.76, SD 5.11). The number needed to treat and 95% CI were calculated to indicate the average number of participants who need to be treated to achieve an additional response compared with the control group [61].

Mediation Analyses

Four mediation analyses were conducted using the PROCESS macro (version 4.0) for SPSS [62]. The models build up on each other to explore their individual and shared contribution in stress reduction. In all models, the independent variable (X) was the study condition, and the dependent variable (Y) was perceived stress (PSS-10 at T3). The proposed mediators were occupational self-efficacy at T2 (PROCESS model 4); occupational self-efficacy at T2 and efforts at T3 (PROCESS model 6); occupational self-efficacy at T2 and rewards at T3 (PROCESS model 6); and occupational self-efficacy, efforts, and rewards at T3 (PROCESS model 81). Baseline scores of the outcome and mediator were considered covariates. For indirect effects that were considered significant if P<.05 and 95% CIs did not cover 0, 10,000 bias-corrected bootstrap samples were applied [62]. An additional sensitivity analysis including only study completers was performed.

Ethics Approval

The Ethical Committee of the Leuphana University of Lueneburg approved the study (reference Ebert201408_Stresstraining). The trial was registered in the German Clinical Trials Register (DRKS00005990).

Results

Participants and Baseline Characteristics

The sample initially consisted of 264 participants of which 2 requested the deletion of assessed data after trial conduction. Consequently, the final sample included 262 participants (182/262, 69.4% female) aged 20-65 years (mean 42.2 years, SD 9.76 years), allocated to either the IG (n=130) or the WLC (n=132) group. Figure 1 depicts the study flow and Table 1 summarizes detailed baseline characteristics. A multivariate ANOVA indicated there was no meaningful difference in baseline outcomes between groups ($F_{19,232}$=1.08, P=.37). Primary outcome data were missing for 9.9% (n=26) at T2 and 15.3% (n=40) at T3. The Little missing completely at random test failed significance, indicating that the null hypothesis proposing patterns of missing values being not dependent on observed and unobserved factors among the participants' values need not be rejected.

Figure 1. Participant flow.

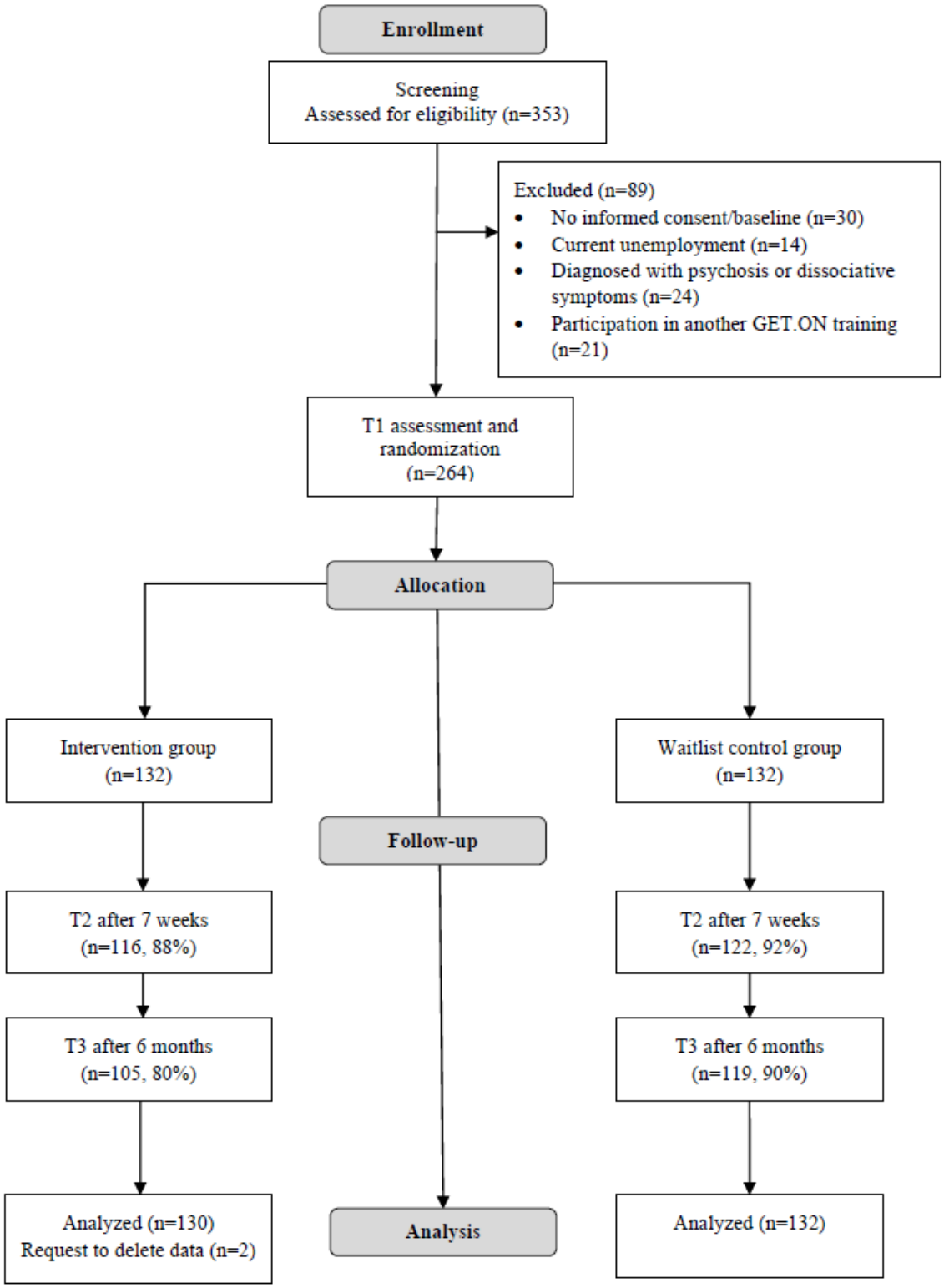


Table 1. Baseline characteristics[a].

Characteristics	All participants (N=262)	IG[b] (n=130)	WLC[c] group (n=132)
Sociodemographic			
Age, mean (SD)	42.20 (9.76)	42.87 (9.54)	43.42 (10.02)
Men, n (%)	80 (30.5)	45 (34.60)	35 (26.5)
Women, n (%)	182 (69.5)	85 (65.40)	97 (73.5)
Diverse, n (%)	N/A[d]	N/A	N/A
Marital status, n (%)			
Single	78 (29.8)	41 (31.5)	37 (28.0)
Married	123 (46.9)	59 (45.4)	64 (48.5)
Cohabited	29 (11.1)	16 (12.3)	13 (9.8)
Divorced	31 (11.8)	14 (10.8)	17 (12.9)
Widowed	1 (0.4)	N/A	1 (0.8)
Educational level, n (%)			
Low	6 (2.3)	2 (1.5)	4 (3.0)
Middle	54 (20.6)	22 (16.9)	32 (24.2)
High	202 (77.1)	106 (81.5)	96 (72.7)
Employment			
Full-time, n (%)	205 (78.2)	105 (80.8)	100 (75.8)
Part-time, n (%)	53 (20.2)	23 (17.7)	30 (22.7)
Sick leave, n (%)	4 (1.5)	2 (1.5)	2 (1.5)
Managerial position, n (%)	100 (38.2)	50 (38.5)	50 (37.9)
Work experience in years, mean (SD)	18.40 (10.83)	17.79 (10.86)	19.01 (10.81)
Work sectors, n (%)			
Service	62 (23.7)	26 (20)	36 (27.3)
Economy	56 (21.4)	22 (16.9)	34 (25.8)
Health	33 (12.6)	22 (16.9)	11 (8.3)
Social	44 (16.8)	19 (14.6)	25 (18.9)
Information technologies	24 (9.2)	15 (11.5)	9 (6.8)
Other	36 (13.7)	19 (14.6)	17 (12.9)
Income, n (%)			
Low	73 (27.9)	29 (22.3)	44 (33.3)
Middle	45 (17.2)	26 (20)	19 (14.4)
High	123 (46.9)	64 (49.2)	59 (44.7)
Use of health services, n (%)			
Previous or current psychotherapy	119 (45.4)	55 (42.3)	64 (48.5)
Experience in health trainings	38 (14.5)	15 (11.5)	23 (17.4)

[a]Values presented only for participants who provided the respective data.

[b]IG: intervention group.

[c]WLC: waitlist control.

[d]N/A: not applicable.

Primary Outcome Measure

ANCOVAs to detect differences between the IG and the WLC group at T2 and T3 revealed significantly lower stress levels assessed with the PSS-10 for the IG at T2 ($F_{259,1}$=46.14, P<.001, *d*=0.87, 95% CI 0.61-1.12, Δ5.00) and T3 ($F_{259,1}$=24.82, P<.001, *d*=0.65, 95% CI 0.41-0.90, Δ4.19). The per-protocol analysis corroborated those results with significant between-group differences at T2 ($F_{173,2}$=34.86, P<.001, *d*=1.04, 95% CI 0.69-1.40, Δ5.79) and T3 ($F_{173,2}$=20.15, P<.001, *d*=0.56, 95% CI 0.22-0.90, Δ3.68). For all outcome measures at T2 and T3, Table 2 displays the means and SDs and Table 3 shows ANCOVA results.

Table 2. Means and SDs of outcome variables at baseline (T1), 7 weeks (T2), and 6 months (T3) after the intervention.

Outcome	T1		T2[a]		T3[a]	
	IG[b]	WLC[c]	IG	WLC	IG	WLC
Primary outcome measure						
Perceived stress	23.76 (5.11)	24.81 (5.03)	18.33 (6.18)	23.33 (5.32)	17.53 (6.42)	21.72 (6.39)
Perceived stress (per-protocol analysis)	23.89 (5.63)	24.79 (5.04)	17.5 (6.14)	23.29 (5.32)	18 (6.97)	21.68 (6.39)
Secondary outcome measures						
Mental health and work related						
Quality of life	0.58 (0.15)	0.55 (0.13)	N/A[d]	N/A	0.68 (0.17)	0.57 (0.17)
Depression	17.05 (6.09)	18.05 (6.43)	13.68 (7.41)	15.76 (7.43)	11.56 (6.74)	14.8 (8.46)
Resilience	20.12 (6.67)	20.29 (6.37)	22.5 (6.08)	19.38 (6.12)	N/A	N/A
Emotional exhaustion	4.57 (0.78)	4.62 (0.73)	4.05 (0.89)	4.52 (0.81)	3.87 (0.94)	4.42 (0.95)
Occupational self-efficacy	22.06 (6.14)	21.58 (6.24)	24.38 (5.41)	22.2 (6.08)	N/A	N/A
Work engagement (vigor)	2.95 (1.21)	2.99 (1.18)	3.06 (1.17)	2.72 (1.15)	3.1 (1.18)	2.75 (1.21)
Work engagement (dedication)	3.31 (1.3)	3.28 (1.37)	3.36 (1.17)	2.98 (1.36)	3.37 (1.27)	3.02 (1.3)
Work engagement (absorption)	3.01 (1.39)	3.01 (1.51)	3.14 (1.28)	2.84 (1.38)	3.14 (1.27)	2.85 (1.38)
Work ability index	5.92 (1.96)	5.86 (1.96)	6.55 (1.88)	5.83 (2.08)	N/A	N/A
Presenteeism	5.01 (2.25)	5.27 (2.58)	4.5 (2.37)	4.94 (2.39)	N/A	N/A
Effort-reward imbalance						
Efforts	10.67 (1.42)	10.5 (1.52)	10.01 (1.76)	10.11 (1.76)	9.72 (1.76)	9.89 (1.83)
Rewards	16.29 (3.74)	15.77 (3.86)	16.69 (3.8)	15.61 (3.96)	17.06 (3.62)	16.03 (3.9)
Ratio	1.62 (0.49)	1.64 (0.49)	1.5 (0.55)	1.62 (0.56)	1.43 (0.47)	1.55 (0.54)

[a]Missing data handled by multiple imputation.

[b]IG: intervention group.

[c]WLC: waitlist control.

[d]N/A: not applicable.

Table 3. Between-group differences at 7 weeks (T2) and 6 months (T3) after the intervention.

Outcomes	T2[a]		T3[a]	
	d (95% CI)	ANCOVA[b] ($F_{259,1}$)	*d* (95% CI)	ANCOVA ($F_{259,1}$)
Primary outcome measure				
Perceived stress	0.87 (0.61 to 1.12)	46.14[c]	0.65 (0.41 to 0.90)	24.82[c]
Perceived stress (per-protocol analysis)[d]	1.04 (0.69 to 1.40)	34.86[c]	0.56 (0.22 to 0.90)	20.15[c]
Secondary outcome measures				
Mental health and work related				
Quality of life[e]	N/A[f]	N/A	0.65 (0.37 to 0.93)	14.44[c]
Depression	0.28 (–0.04 to 0.52)	3.55	0.42 (0.18 to 0.67)	9.99[g]
Resilience	N/A	N/A	0.51 (0.26 to 0.76)	31.72[c]
Emotional exhaustion	0.56 (0.31 to 0.80)	25.36[c]	0.59 (0.34 to 0.83)	25.79[c]
Occupational self-efficacy	0.38 (0.13 to 0.62)	10.65[g]	N/A	N/A
Work engagement (vigor)	0.29 (0.05 to 0.53)	9.30[g]	0.29 (0.05 to 0.53)	7.61[g]
Work engagement (dedication)	0.30 (0.06 to 0.54)	8.71[g]	0.27 (0.03 to 0.52)	6.22[h]
Work engagement (absorption)	0.41 (0.66 to 0.17)	5.80[h]	0.22 (0.03 to 0.46)	3.97[h]
Presenteeism	0.18 (–0.06 to 0.43)	1.61	N/A	N/A
Work ability index	0.36 (0.12 to 0.60)	9.19[g]	N/A	N/A
Effort-reward imbalance				
Efforts	0.05 (–0.19 to 0.30)	1.72	0.09 (–0.15 to 0.34)	1.87
Rewards	0.28 (0.04 to 0.52)	4.42[h]	0.27 (–0.03 to 0.52)	3.72
Ratio	0.22 (0.03 to 0.50)	4.21[h]	0.24 (0.01 to 0.48)	4.07[h]

[a]Missing data handled by multiple imputation.

[b]ANCOVA: analysis of covariance

[c]Significance level used: P<.001.

[d]$F_{173,2}$.

[e]$F_{200,1}$.

[f]N/A: not applicable.

[g]Significance level used: P<.01.

[h]Significance level used: P<.05.

Response Analyses

At T2, significantly more participants in the IG (65/130, 50%) showed a reliable improvement in perceived stress measured with the PSS-10 compared with the WLC group (33/132, 25%) and significantly fewer participants in the IG (4/130, 3.1%) experienced symptom deterioration compared with the WLC group (14/132, 10.6%; χ^2_2=19.94, P<.001). The number needed to treat to achieve reliable improvement was 4 (95% CI 2.8-7.3). The number of symptom-free participants at T2 was significantly higher in the IG (39/130, 30%) compared with the WLC group (7/132, 5.3%; χ^2_1=23.52, P<.001).

Secondary Outcome Measures

The ANCOVAs showed significant between-group differences for most secondary outcome measures (Table 3). Positive impacts for participants in the IG compared with the WLC group were found at T2 and T3 for occupational self-efficacy (measured with the OSS-SF), burnout (assessed with the MBI-GS-D), work engagement (assessed with the UWES), and work ability (Work Ability Index). Effect sizes (*d*) ranged from 0.29 (95% CI 0.05-0.53; UWES scale vigor) to 0.56 (95% CI 0.31-0.80; MBI-GS-D) at T2 and from 0.22 (95% CI 0.03-0.46; UWES scale absorption) to 0.65 (95% CI 0.37-0.93; AQoL) at T3. Scores between groups did not significantly differ for depression (CES-D; *d*=0.28, 95% CI –0.04 to 0.52, P=.06) and work limitations (Work Limitations Questionnaire; *d*=0.18, 95% CI –0.06 to 0.43, P=.21) at T2. Regarding the effort-reward imbalance, participants in the IG showed significantly higher values for rewards at T2 (*d*=0.28, 95% CI 0.04-0.52, P=.04), whereas between-group scores did not significantly differ for efforts at T2 (*d*=0.05, 95% CI –0.19 to 0.30, P=.19) and T3

(d=0.09, 95% CI –0.15 to 0.34, P=.17), and for rewards at T3 (d=0.27, 95% CI –0.03 to 0.52, P=.06).

Mediation Analyses

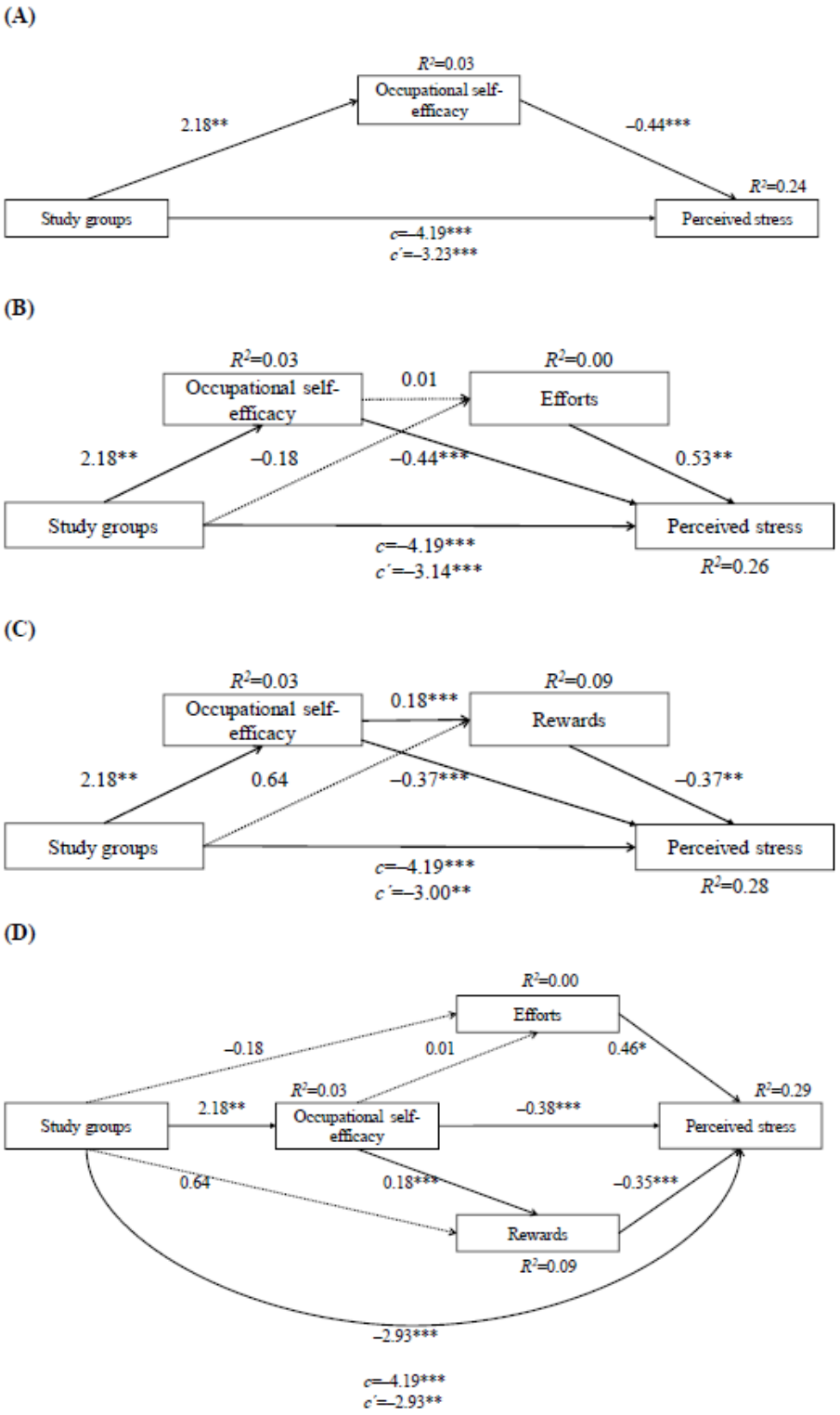
Figure 2 depicts the 4 mediation analyses performed. Results of the first model (Figure 2A) showed that the unstandardized regression coefficient for the study groups (X) predicting stress (Y) was significant (c=–4.19, t_{260}=–5.29, P<.001). Occupational self-efficacy (M) was found to be a significant mediator for this effect (b=–0.44, t_{258}=–6.87, P<.001). Furthermore, the study group had a significant effect on occupational self-efficacy (b=2.18, t_{260}=3.06, P<.002). The direct effect remained significant after incorporating the mediating variable into the model (c'=–3.23, t_{260}=–4.36, P<.001). The indirect effect was significant (b=0.95, 95% CI –1.73 to –0.32, P<.001). This model accounted for 24% of the variance (R^2) in stress reduction.

The second mediation model (Figure 2B) with occupational self-efficacy as M_1 and efforts as M_2 revealed significant total (c=–4.19, t_{260}=–5.29, P<.001) and direct (c'=–3.14, t_{260}=–4.36, P<.001) effects. Occupational self-efficacy (M_1) significantly mediated the effect on stress (b=–0.44, t_{258}=–6.59, P<.001), whereas it had no significant effect on efforts (M_2; b=0.01, t_{259}=0.27, P=.79). However, efforts (M_2) were significantly associated with stress (b=0.53, t_{258}=2.64, P=.008). The study group had no significant effect on efforts (M_2; b=–0.18, t_{258}=–0.79, P=.43). Therefore, a significant indirect mediating effect was only found for the association of occupational self-efficacy with the study group (b=–0.14, 95% CI –0.25 to –0.05, P<.001). Together, 26% of the variance (R^2) in perceived stress was explained.

After incorporating occupational self-efficacy as M_1 and rewards as M_2, the mediation model (Figure 2C) resulted in significant total (c=–4.19, t_{260}=–5.29, P<.001) and direct (c'=–3.00, t_{260}=–4.12, P<.001) effects. Occupational self-efficacy (M_1) significantly mediated the effect on stress (b=–0.37, t_{258}=–5.76, P<.001) and rewards (M_2; b=0.18, t_{259}=4.52, P<.001). Rewards (M_2) could significantly predict stress (b=–0.37, t_{258}=–3.79, P<.001). Comparable to the preceding mediation model, a significant indirect mediation effect for the association between the intervention and stress as an outcome could be found for occupational self-efficacy (b=–0.12, 95% CI –0.22 to –0.04, P<.001). Furthermore, the indirect path taking occupational self-efficacy (M_1) and rewards (M_2) between the study group and perceived stress into account was significant (b=–0.02, 95% CI –0.05 to –0.01, P<.001). Participation in the intervention did not significantly predict rewards (M_2; P=.16). In total, all variables accounted for R^2=0.28.

The fourth mediation model (Figure 2D) that incorporated all mediators (M_1: occupational self-efficacy, M_2: efforts, and M_3: rewards) again resulted in significant total (c=–4.19, t_{260}=–5.29, P<.001) and direct (c'=–2.93, t_{260}=–4.06, P<.001) effects. Occupational self-efficacy (M_1) significantly predicted perceived stress (b=–0.38, t_{257}=–5.91, P<.001) and rewards (M_3) (b=0.18, t_{259}=4.52, P<.001), yet not efforts (M_2; b=0.01, t_{259}=0.27, P=.79). The effect on stress was also significantly predicted by both efforts (M_2; b=0.46, t_{257}=2.32, P=.02) and rewards (M_3; b=–0.35, t_{257}=–3.56, P<.001). The study group did not significantly predict neither efforts (M_2) nor rewards (M_3) directly. Altogether, significant indirect paths between the study group and perceived stress were found for occupational self-efficacy (M_1; b=–0.12, 95% CI –0.22 to –0.04, P<.001) as well as for occupational self-efficacy (M_1) and rewards (M_3; b=–0.02, 95% CI –0.05 to –0.01, P<.001). This final model including all proposed mediators together explained 29% of the variance (R^2) in stress reduction. For all models, sensitivity analyses performed including only study completers corroborated the results.

Figure 2. Mediation analyses with study condition as independent variable (X) and perceived stress (PSS-10) at T3 as dependent variable (Y) for all models. Proposed mediators: (A) Occupational self-efficacy at T2; (B) Occupational self-efficacy at T2 and efforts at T3; (C) Occupational self-efficacy at T2 and rewards at T3; and (D) Occupational self-efficacy at T2, and efforts and rewards at T3. Study conditions are coded 0=wait list control group, 1=intervention group. The figure includes unstandardized β coefficients and illustrates significant (solid line) and non-significant (dotted line) effects between variables, total (c) and direct (c′) effects. Significance levels used: ***P<.001, **P<.01, *P<.05.



Discussion

Principal Findings

Results of this study confirm that the SMI could effectively reduce stress in employees perceiving elevated stress levels and even when they were exposed to a high load of efforts that is not adequately balanced by rewards. Secondary analyses demonstrated the beneficial effects for mental health and work-related outcomes as well as for rewards. Step-by-step mediation analyses revealed that the participation in the intervention significantly predicted occupational self-efficacy,

which describes the confidence of an individual to handle any challenges at work and which was a mediator in the effect on stress and rewards that again predicted stress. All 3 investigated mediators (ie, occupational self-efficacy, efforts, and rewards) were significantly associated with perceived stress. However, neither participation in the SMI nor the increase in occupational self-efficacy enabled employees to achieve favorable effects on the level of efforts, while efforts still enfolded an adverse effect on perceived stress.

The results revealed practically meaningful effect sizes for stress reduction. A similar effect was found in another trial on the same SMI with adherence-focused guidance [45] and our study extend those results by the inclusion of a high-risk population that experiences adverse working conditions. Compared with a study on the same SMI with more intensive guidance [22], the effect sizes were not as large at follow-up. This raises the question as to whether more personal support from a mental health expert, which is expected to be conducive to the efficacy of an SMI [19,22], might aid participants that experience greater difficulties in their stress management due to adverse working conditions. Considering the efficacy of occupational web-based interventions in general, results from this study are in line with demonstrated average effect sizes in a recent meta-analysis [19] and revealed significant improvements in a variety of outcomes on mental health and work-related levels. For example, participants in the IG showed lower levels of emotional exhaustion, more resilience, and higher work engagement, as well as vigor, dedication, and absorption at work. No significant between-group effects were found for presenteeism, while mixed results were obtained for depression. The detected effect sizes for engagement and presenteeism compare with a recent meta-analysis for occupational web-based interventions [63]. Moreover, the participation significantly increased occupational self-efficacy that was shown to be a relevant mediator in the efficacy of the SMI on stress reduction. These results support findings of another RCT on the same SMI showing significant effects on occupational self-efficacy [34] and positive associations between stress levels and self-efficacy [30]. The obtained results for the effort-reward imbalance tie well with mixed effects found in studies on the same SMI for the effort-reward imbalance ratio [34] and for efforts and rewards evaluated as separate outcomes [45], demonstrating that web-based SMIs enfold substantially larger effects on individuals' health compared with perceived working conditions and organizational characteristics [17].

To examine whether and how an increase in personal resources could support participants in achieving successful stress reduction despite facing adverse working conditions, mediating effects were investigated not only for occupational self-efficacy, but also for efforts and rewards of the workplace. The 4 mediation analyses conducted progressively accounted for the variance in perceived stress. The first model (Figure 2A) confirmed that the participation in the intervention successfully increased occupational self-efficacy, which in turn had a significant effect on stress reduction. This is in line with evidence showing that higher levels of self-efficacy are associated with lower levels of work stress and the assumption that problem-solving skills increase the confidence of an individual to be able to proactively reduce stressors and increase rewarding situations [30]. The second mediation model (Figure 2B) showed that the intervention's positive effect on occupational self-efficacy did not affect efforts that were negatively associated with stress. This is in line with another SMI study on teachers which showed that participants could influence rewards, yet not efforts [64]. One potential reason for the lack of association could be the design of the intervention that did not predefine the topics participants should reflect on in the problem-solving exercises and if the focus was on job-related or personal stressors. Furthermore, this portrays one of the core premises of the effort-reward imbalance model [6], that is, an increased degree of efforts necessary to spend at work is associated with high strain. The third mediation analysis (Figure 2C) revealed a significant relationship between participation in the SMI, occupational self-efficacy, and rewards. This is in line with evidence showing that occupational self-efficacy is substantially associated with affective commitment that might motivate employees to increase their job resources within their company [65]. Comparable to the precedent mediation model, rewards were significantly associated with stress, which is in line with the effort-reward imbalance model [6]. The final mediation analysis (Figure 2D) incorporated the 3 models. Occupational self-efficacy was significantly increased and a mediator in the relationship between the study group and outcome. Although both efforts and rewards predicted levels of stress, the intervention only had an impact on rewards, but not on efforts, with occupational self-efficacy seemingly playing a mediating role in this association. However, both efforts and rewards had significant effects on stress.

Limitations

Several limitations should be considered. Despite the positive effects of the individual-focused intervention on employees' mental health, the persisting adverse effects of efforts indicate that this approach might be incomplete. Therefore, it should be investigated whether a combination of individual- and organizational-focused digital interventions will contribute to more comprehensive effects on employees' mental health [66]. Positive effects of occupational self-efficacy in individual-focused interventions might help employees to engage more confidently in organizational-focused interventions. Furthermore, the generalizability of the results might be limited. In contrast to recruitment on a company level, the applied open recruitment strategy addressed participants directly, which was shown to be associated with effects on personal health outcomes for occupational SMIs [19]. In this study, participants in the IG received adherence-focused guidance that was established and shown to be effective in previous studies [45,67]. Given the notion that guidance is supposed to be conducive to the efficacy of SMIs [19,68] and its low intensity in the adherence-focused format, further research could investigate whether a higher intensity in guidance might facilitate the efficacy of the SMI for participants that experience greater difficulties for successful changes due to adverse workplace conditions. Concerning the mediators, a methodological limitation might be the selection of measures in this study because participants might have been encouraged to make changes to aspects of their work that were

not captured in this trial (eg, conflict between work and private life) [69]. Despite this, this trial provides valuable first insights into if and how a web-based SMI can be effective within a high-risk population despite their exposure to adverse working conditions.

Conclusion and Practical Implications

To conclude, this trial aimed to expand research on the efficacy of web-based SMIs and to add valuable insights into the scarce evidence for high-risk populations. To the best of our knowledge, this is the first trial demonstrating positive effects of a web-based SMI on stress reduction in employees despite their adverse working conditions. In-depth analyses examining mechanisms of change suggest that the SMI increased occupational self-efficacy that mediated the intervention's effect on stress. Furthermore, both efforts and rewards predicted levels of stress, yet the intervention only had an impact on rewards, with occupational self-efficacy seemingly playing a mediating role in this association. It seems vital to note that this web-based intervention could improve health at work within a short period and without any direct changes to working conditions. Further medium- and long-term improvements would be possible if complex organizational interventions were introduced to reduce stressors in the workplace. For practice, these results have several implications. First, the implementation of the web-based SMI can be recommended due to its beneficial health effects even if employees experience adverse working conditions. Second, occupational self-efficacy should be considered as an important concept in the design of an SMI. Third, the limited effects of the SMI on the perception of working conditions underline that organizational top-down changes are still indispensable. Future studies could further investigate which factors contribute to the efficacy of a person-centered intervention on working conditions and examine, for example, the role of guidance.

Acknowledgments

The German health care insurance company BARMER and the European Commission funded this study (EFRE: ZW6-80119999, CCI 2007DE161PR001).

Conflicts of Interest

DDE is a stakeholder in the Institute for Online Health Training that aims to transfer scientific knowledge related to this research into routine health care.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1220 KB - jmir_v24i10e40488_app1.pdf]

References

1. World Health Organization (WHO). The Global Burden of Disease: 2004 Update. Geneva, Switzerland: WHO; 2008.
2. Niedhammer I, Bertrais S, Witt K. Psychosocial work exposures and health outcomes: a meta-review of 72 literature reviews with meta-analysis. Scand J Work Environ Health 2021 Oct 01;47(7):489-508 [FREE Full text] [doi: 10.5271/sjweh.3968] [Medline: 34042163]
3. Niedhammer I, Sultan-Taïeb H, Parent-Thirion A, Chastang J. Update of the fractions of cardiovascular diseases and mental disorders attributable to psychosocial work factors in Europe. Int Arch Occup Environ Health 2022 Jan 28;95(1):233-247 [FREE Full text] [doi: 10.1007/s00420-021-01737-4] [Medline: 34181059]
4. Theorell T, Hammarström A, Aronsson G, Träskman Bendz L, Grape T, Hogstedt C, et al. A systematic review including meta-analysis of work environment and depressive symptoms. BMC Public Health 2015 Aug 01;15:738 [FREE Full text] [doi: 10.1186/s12889-015-1954-4] [Medline: 26232123]
5. Hassard J, Teoh KRH, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: A systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
6. Siegrist J. Adverse health effects of high-effort/low-reward conditions. Journal of Occupational Health Psychology 1996;1(1):27-41. [doi: 10.1037/1076-8998.1.1.27]
7. Rugulies R, Aust B, Madsen IE. Effort-reward imbalance at work and risk of depressive disorders. A systematic review and meta-analysis of prospective cohort studies. Scand J Work Environ Health 2017 Jul 01;43(4):294-306 [FREE Full text] [doi: 10.5271/sjweh.3632] [Medline: 28306759]
8. Eddy P, Heckenberg R, Wertheim EH, Kent S, Wright BJ. A systematic review and meta-analysis of the effort-reward imbalance model of workplace stress with indicators of immune function. J Psychosom Res 2016 Dec;91:1-8. [doi: 10.1016/j.jpsychores.2016.10.003] [Medline: 27894456]
9. Dragano N, Siegrist J, Nyberg ST, Lunau T, Fransson EI, Alfredsson L, IPD-Work consortium. Effort-Reward Imbalance at Work and Incident Coronary Heart Disease: A Multicohort Study of 90,164 Individuals. Epidemiology 2017 Jul;28(4):619-626 [FREE Full text] [doi: 10.1097/EDE.0000000000000666] [Medline: 28570388]

10. Siegrist J, Li J. Associations of Extrinsic and Intrinsic Components of Work Stress with Health: A Systematic Review of Evidence on the Effort-Reward Imbalance Model. Int J Environ Res Public Health 2016 Apr 19;13(4):432 [FREE Full text] [doi: 10.3390/ijerph13040432] [Medline: 27104548]
11. Tsutsumi A, Kawakami N. A review of empirical studies on the model of effort-reward imbalance at work: reducing occupational stress by implementing a new theory. Soc Sci Med 2004 Dec;59(11):2335-2359. [doi: 10.1016/j.socscimed.2004.03.030] [Medline: 15450708]
12. van Vegchel N, de Jonge J, Bosma H, Schaufeli W. Reviewing the effort-reward imbalance model: drawing up the balance of 45 empirical studies. Soc Sci Med 2005 Mar;60(5):1117-1131. [doi: 10.1016/j.socscimed.2004.06.043] [Medline: 15589679]
13. EU-OSHA. Expert forecast on emerging psychosocial risks related to occupational safety and health. European Risk Observatory Report. Luxembourg City, Luxembourg: Office for Official Publications of the European Communities; 2007.
14. Canadian Standards Association, Bureau de Normalisation. Psychological health and safety in the workplace ? Prevention, promotion, and guidance to staged implementation (CSA Publication No. CAN/CSAZ1003-13/BNQ9700-803/2013). Toronto, ON, Canada: CSA Group and BNQ; 2013.
15. Dahl MS. Organizational Change and Employee Stress. Management Science 2011 Feb;57(2):240-256. [doi: 10.1287/mnsc.1100.1273]
16. Carolan S, Harris PR, Cavanagh K. Improving Employee Well-Being and Effectiveness: Systematic Review and Meta-Analysis of Web-Based Psychological Interventions Delivered in the Workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
17. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
18. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
19. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 01;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]
20. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
21. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
22. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
23. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The Implementation of Internet Interventions for Depression: A Scoping Review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
24. Behrendt D, Ebert DD, Spiegelhalder K, Lehr D. Efficacy of a Self-Help Web-Based Recovery Training in Improving Sleep in Workers: Randomized Controlled Trial in the General Working Population. J Med Internet Res 2020 Jan 07;22(1):e13346 [FREE Full text] [doi: 10.2196/13346] [Medline: 31909725]
25. Geraedts AS, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P. Short-term effects of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 May 06;16(5):e121 [FREE Full text] [doi: 10.2196/jmir.3185] [Medline: 24800966]
26. Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, van Mechelen W, Cuijpers P. Long-term results of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 Jul 09;16(7):e168 [FREE Full text] [doi: 10.2196/jmir.3539] [Medline: 25008127]
27. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York, NY: Springer; 1984.
28. Tims MB, Bakker A, Derks D. Daily job crafting and the self-efficacy – performance relationship. Journal of Managerial Psychology 2014;29(5):490-507 [FREE Full text] [doi: 10.1108/JMP-05-2012-0148]
29. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Advances in Behaviour Research and Therapy 1978 Jan;1(4):139-161. [doi: 10.1016/0146-6402(78)90002-4]
30. Jex SM, Bliese PD, Buzzell S, Primeau J. The impact of self-efficacy on stressor-strain relations: coping style as an explanatory mechanism. J Appl Psychol 2001 Jun;86(3):401-409. [doi: 10.1037/0021-9010.86.3.401] [Medline: 11419800]
31. Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. Scand J Work Environ Health 2014 Nov 13;40(6):582-596 [FREE Full text] [doi: 10.5271/sjweh.3449] [Medline: 25121986]

32. Rigotti T, Schyns B, Mohr G. A Short Version of the Occupational Self-Efficacy Scale: Structural and Construct Validity Across Five Countries. Journal of Career Assessment 2008 May 01;16(2):238-255. [doi: 10.1177/1069072707305763]
33. Guarnaccia C, Scrima F, Civilleri A, Salerno L. The Role of Occupational Self-Efficacy in Mediating the Effect of Job Insecurity on Work Engagement, Satisfaction and General Health. Curr Psychol 2016 Nov 12;37(3):488-497. [doi: 10.1007/s12144-016-9525-0]
34. Ebert DD, Franke M, Zarski A, Berking M, Riper H, Cuijpers P, et al. Effectiveness and Moderators of an Internet-Based Mobile-Supported Stress Management Intervention as a Universal Prevention Approach: Randomized Controlled Trial. J Med Internet Res 2021 Dec 22;23(12):e22107 [FREE Full text] [doi: 10.2196/22107] [Medline: 34941541]
35. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Sep 01;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
36. Siegrist J. Effort-reward imbalance at work and health. In: Perrewe PL, Ganster DC, editors. Historical and Current Perspectives on Stress and Health (Research in Occupational Stress and Well Being, Vol. 2). Bingley, UK: Emerald Group Publishing Limited; 2002:261-291.
37. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ 1996 Sep 07;313(7057):570-571 [FREE Full text] [doi: 10.1136/bmj.313.7057.570] [Medline: 8806240]
38. 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]
39. Michalsen A, Jeitler M, Brunnhuber S, Lüdtke R, Büssing A, Musial F, et al. Iyengar yoga for distressed women: a 3-armed randomized controlled trial. Evid Based Complement Alternat Med 2012;2012:408727-408723 [FREE Full text] [doi: 10.1155/2012/408727] [Medline: 23049608]
40. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
41. Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: effort-reward imbalance. Int Arch Occup Environ Health 2009 Aug 19;82(8):1005-1013. [doi: 10.1007/s00420-008-0384-3] [Medline: 19018554]
42. Lehr D, Koch S, Hillert A. Where is (im)balance? Necessity and construction of evaluated cut-off points for effort-reward imbalance and overcommitment. Journal of Occupational and Organizational Psychology 2010 Mar;83(1):251-261. [doi: 10.1348/096317909X406772]
43. Beck AT, Steer RA, Brown GK. Beck Depression Inventory Manual. 2nd ed. San Antonio, TX: Psychological Corporation; 1996.
44. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May 16;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
45. Nixon P, Boß L, Heber E, Ebert DD, Lehr D. A three-armed randomised controlled trial investigating the comparative impact of guidance on the efficacy of a web-based stress management intervention and health impairing and promoting mechanisms of prevention. BMC Public Health 2021 Aug 05;21(1):1511 [FREE Full text] [doi: 10.1186/s12889-021-11504-2] [Medline: 34353294]
46. Schaufeli W, Leiter MP, Maslach C, Jackson SE. Maslach Burnout Inventory - General Survey (MI-GS). In: Maslach C, Jackson SE, Leiter MP, editors. Maslach Burnout Inventory Manual. Maslach Burnout Inventory Manual Palo Alto, CA: Consulting Psychologists Press; Jan 1996:e1.
47. Schaufeli W, Salanova M, González-Romá V, Bakker A. The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. Journal of Happiness Studies 2002 Feb;3(1):71-92. [doi: 10.1023/A:1015630930326]
48. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health--a prospective study of women on long-term sick leave. Scand J Work Environ Health 2010 Sep 07;36(5):404-412 [FREE Full text] [doi: 10.5271/sjweh.2917] [Medline: 20372766]
49. Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care 2001 Jan;39(1):72-85. [doi: 10.1097/00005650-200101000-00009] [Medline: 11176545]
50. Hautzinger M. Allgemeine Depressions Skala (Manual). Psycontent. Göttingen, Germany: Beltz Test GmbH; 1993. URL: http://www.psycontent.com/content/3x081j2m2732r345/ [accessed 2022-09-21]
51. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
52. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. J Trauma Stress 2007 Dec;20(6):1019-1028. [doi: 10.1002/jts.20271] [Medline: 18157881]
53. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient 2014 Nov 23;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]

54. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and Validity of Assessing User Satisfaction With Web-Based Health Interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
55. IBM Corp.. SPSS Statistics for Windows. IBM Corp. Armonk, NY: IBM Corp; 2017. URL: https://www.ibm.com/in-en/products/spss-statistics [accessed 2022-09-21]
56. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. BMC Med Res Methodol 2014 Apr 09;14:49 [FREE Full text] [doi: 10.1186/1471-2288-14-49] [Medline: 24712304]
57. O'Connell NS, Dai L, Jiang Y, Speiser JL, Ward R, Wei W, et al. Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods. J Biom Biostat 2017 Feb 24;8(1):1-8 [FREE Full text] [doi: 10.4172/2155-6180.1000334] [Medline: 30555734]
58. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods 2002;7(2):147-177. [doi: 10.1037/1082-989x.7.2.147]
59. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]
60. Janicki-Deverts D, Cohen S. Who's Stressed? Distributions of Psychological Stress in the United States in Probability Samples from 1983, 2006, and 2009. Journal of Applied Social Psychology 2012 Jun;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]
61. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ 1999 Dec 04;319(7223):1492-1495 [FREE Full text] [doi: 10.1136/bmj.319.7223.1492] [Medline: 10582940]
62. Hayes A. Introduction to Mediation, Moderation and Conditional Process Analysis. A Regression-Based Approach (2nd Edition). New York, NY: The Guilford Press; 2018.
63. Stratton E, Jones N, Peters S, Torous J, Glozier N. Digital mHealth Interventions for Employees: Systematic Review and Meta-Analysis of Their Effects on Workplace Outcomes. J Occup Environ Med 2021 Aug 01;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
64. Unterbrink T, Pfeifer R, Krippeit L, Zimmermann L, Rose U, Joos A, et al. Burnout and effort-reward imbalance improvement for teachers by a manual-based group program. Int Arch Occup Environ Health 2012 Aug;85(6):667-674. [doi: 10.1007/s00420-011-0712-x] [Medline: 22038086]
65. Schyns B, von Collani G. A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. European Journal of Work and Organizational Psychology 2002 Jun;11(2):219-241. [doi: 10.1080/13594320244000148]
66. Engels M, Boß L, Engels J, Kuhlmann R, Kuske J, Lepper S, et al. Facilitating stress prevention in micro and small-sized enterprises: protocol for a mixed method study to evaluate the effectiveness and implementation process of targeted web-based interventions. BMC Public Health 2022 Mar 26;22(1):591 [FREE Full text] [doi: 10.1186/s12889-022-12921-7] [Medline: 35346127]
67. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-Based Mobile-Supported Stress Management: A Pooled Analysis of Individual Participant Data From Three Randomized Controlled Trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
68. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
69. Clausen T, Madsen IE, Christensen KB, Bjorner JB, Poulsen OM, Maltesen T, et al. The Danish Psychosocial Work Environment Questionnaire (DPQ): Development, content, reliability and validity. Scand J Work Environ Health 2019 Jul 01;45(4):356-369 [FREE Full text] [doi: 10.5271/sjweh.3793] [Medline: 30592500]

Abbreviations

ANCOVA: analysis of covariance
AQoL: Assessment of Quality of Life 8D Multi-Attribute Utility Instrument
CES-D: Centre for Epidemiological Studies' Depression Scale
CONSORT: Consolidated Standards of Reporting Trials
IG: intervention group
MBI-GS-D: Maslach Burnout Inventory
N/A: not applicable
OSS-SF: short form of the Occupational Self-Efficacy Scale
PSS-10: 10-item Perceived Stress Scale
RCT: randomized controlled trial
SMI: stress management intervention
UWES: Utrecht Work Engagement Scale
WLC: waitlist control

WLQ: Work Limitations Questionnaire

Edited by T Leung; submitted 23.06.22; peer-reviewed by E Stratton; comments to author 21.07.22; revised version received 11.08.22; accepted 24.08.22; published 20.10.22.

Please cite as:
Nixon P, Ebert DD, Boß L, Angerer P, Dragano N, Lehr D
The Efficacy of a Web-Based Stress Management Intervention for Employees Experiencing Adverse Working Conditions and Occupational Self-efficacy as a Mediator: Randomized Controlled Trial
J Med Internet Res 2022;24(10):e40488
URL: https://www.jmir.org/2022/10/e40488
doi:10.2196/40488
PMID:36264607

Original Paper

Evaluation of a Web-Based Culturally Sensitive Educational Video to Facilitate Informed Cervical Cancer Screening Decisions Among Turkish- and Moroccan-Dutch Women Aged 30 to 60 Years: Randomized Intervention Study

Nora Hamdiui[1,2], PhD; Mart L Stein[1], PhD; Jim van Steenbergen[1,3], PhD; Rik Crutzen[4], Prof Dr; Martine Bouman[5], Prof Dr; Abresham Khan[1], MSc; Miyase N Çetin[1], MSc; Aura Timen[1,6], Prof Dr; Maria van den Muijsenbergh[2,7], Prof Dr

[1]National Coordination Centre for Communicable Disease Control, Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, Netherlands

[2]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[3]Leiden University Medical Centre, Centre for Infectious Diseases, Leiden, Netherlands

[4]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[5]Center for Media & Health, Gouda, Netherlands

[6]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Free University Amsterdam, Amsterdam, Netherlands

[7]Program Prevention and Care, Pharos: Dutch Centre of Expertise on Health Disparities, Utrecht, Netherlands

Corresponding Author:
Nora Hamdiui, PhD
National Coordination Centre for Communicable Disease Control
Centre for Infectious Disease Control
National Institute for Public Health and the Environment
Postbus 1
Bilthoven, 3720 BA
Netherlands
Phone: 31 088 689 7000
Email: nora.hamdiui@rivm.nl

Abstract

Background: In the Netherlands, since 1996, a national cervical cancer (CC) screening program has been implemented for women aged 30 to 60 years. Regional screening organizations send an invitation letter and information brochure in Dutch to the home addresses of targeted women every 5 years. Although this screening is free of charge, Turkish- and Moroccan-Dutch women, especially, show low screening participation and limited informed decision-making (IDM). As Turkish- and Moroccan-Dutch women indicated their need for information on the practical, emotional, cultural, and religious aspects of CC screening, we developed a culturally sensitive educational video (CSEV) as an addition to the current information brochure.

Objective: In this study, we aimed to evaluate the added effect of the CSEV on IDM regarding CC screening participation among Turkish and Moroccan women aged 30 to 60 years in the Netherlands through a randomized intervention study.

Methods: Initial respondents were recruited via several social media platforms and invited to complete a web-based questionnaire. Following respondent-driven sampling, respondents were asked to recruit a number of peers from their social networks to complete the same questionnaire. Respondents were randomly assigned to the control (current information brochure) or intervention condition (brochure and CSEV). We measured respondents' knowledge and attitude regarding CC screening and their intention to participate in the next CC screening round before and after the control or intervention condition. We evaluated the added effect of the CSEV (above the brochure) on their knowledge, attitude, intention, and IDM using intention-to-treat analyses.

Results: The final sample (n=1564) included 686 (43.86%) Turkish and 878 (56.14%) Moroccan-Dutch women. Of this sample, 50.7% (793/1564) were randomized to the control group (350/793, 44.1% Turkish and 443/793, 55.9% Moroccan) and 49.3% (771/1564) to the intervention group (336/771, 43.6% Turkish and 435/771, 56.4% Moroccan). Among the Turkish-Dutch women, 33.1% (116/350) of the control respondents and 40.5% (136/336) of the intervention respondents consulted the brochure (not statistically significant). Among Moroccan-Dutch women, these percentages were 28.2% (125/443) and 37.9% (165/435),

respectively (*P*=.003). Of all intervention respondents, 96.1% (323/336; Turkish) and 84.4% (367/435; Moroccan) consulted the CSEV. The CSEV resulted in more positive screening attitudes among Moroccan-Dutch women than the brochure (323/435, 74.3% vs 303/443, 68.4%; *P*=.07). Women, who had never participated in CC screening before, showed significantly more often a positive attitude toward CC screening compared with the control group (*P*=.01).

Conclusions: Our short and easily implementable CSEV resulted in more positive screening attitudes, especially in Moroccan-Dutch women. As the CSEV was also watched far more often than the current brochure was read, this intervention can contribute to better reach and more informed CC screening decisions among Turkish- and Moroccan-Dutch women.

Trial Registration: International Clinical Trial Registry Platform NL8453; https://tinyurl.com/2dvbjxvc

(J Med Internet Res 2022;24(10):e35962) doi:10.2196/35962

KEYWORDS

cervical cancer; screening; informed decision-making; web-based intervention; culturally sensitive educational video; Turkish; Moroccan; The Netherlands

Introduction

Background

Cervical cancer (CC) is ranked as the fourth most frequently diagnosed cancer in women worldwide [1]. Since the introduction of widespread screening programs, there has been a decline in early- and late-stage CC [2].

In The Netherlands, since 1996, a national CC screening program has been implemented for women aged 30 to 60 years. Regional screening organizations send an invitation letter and information brochure in Dutch to the home addresses of targeted women every 5 years. Screening is free of charge and is carried out by the general practitioner (GP) or their practice assistant who samples a cervical smear (ie, clinician-based sampling). The smear is initially tested for the presence of high-risk human papillomavirus (hrHPV), a risk factor for developing CC [3]. If hrHPV is present, the cervical cells in the smear are assessed for abnormal or precancerous lesions. An important advantage of HPV-based screening is that it can also be performed by self-sampling. If this self-sample tests positive for hrHPV, a cervical smear for cytological examination is sampled at the GP's office.

From an individual's perspective, deciding to participate in screening involves careful consideration of the uncertain benefits and risks of adverse effects. This consideration is pivotal in informed decision-making (IDM), the process in which individuals base their decisions by optimal use of the information and weighing all the aspects involved. IDM is only possible when a woman has adequate decision-relevant knowledge and her attitude toward participating is consistent with her (intended) participation [4].

In the Netherlands, especially Turkish- and Moroccan-Dutch women, representing the largest immigrant population, show low screening participation and limited IDM regarding participation [5,6]. Earlier research indicated an overall lack of knowledge and nonfamiliarity with the possible disadvantages of CC screening [5].

In decision-making, Turkish- and Moroccan-Dutch women consider not only factual medical information but also practical, emotional, cultural, and religious aspects before deciding whether to screen for CC [5]. However, the current invitation letters and information brochures predominantly contain factual medical information. Turkish and Moroccan-Dutch women often indicated not (thoroughly) reading the invitation letter and brochure, or simply being unable to understand these materials due to a lack of good command of the Dutch language [5]. These women were also shown to make less use of printed media and more of audiovisual media [7]. As a culturally competent educational film, which was developed with peer educators, was successful in improving IDM for prenatal screening among pregnant ethnic minority women, we considered this beneficial for IDM in CC screening participation [8]. Thus, we developed a culturally sensitive educational video (CSEV) that incorporates more affective information and distributed it via respondent-driven sampling (RDS).

Objectives

In this study, we evaluated the effect of the CSEV on IDM regarding CC screening participation among Turkish- and Moroccan-Dutch women. We hypothesized that adding a CSEV to the current Dutch information brochure would increase the IDM to participate in CC screening among these women.

Methods

Study Design

Between November 23, 2020, and August 6, 2021, a randomized intervention study was conducted with control and intervention groups. We used web-based RDS to recruit Turkish- and Moroccan-Dutch women, as previous attempts have shown that traditional random sampling methods are not effective in reaching these populations effectively [9]. Their close-knit social networks also enable respondents to recruit each other easily [10]. The reporting of this study adheres to the CONSORT (Consolidated Standards of Reporting Trials) guidelines.

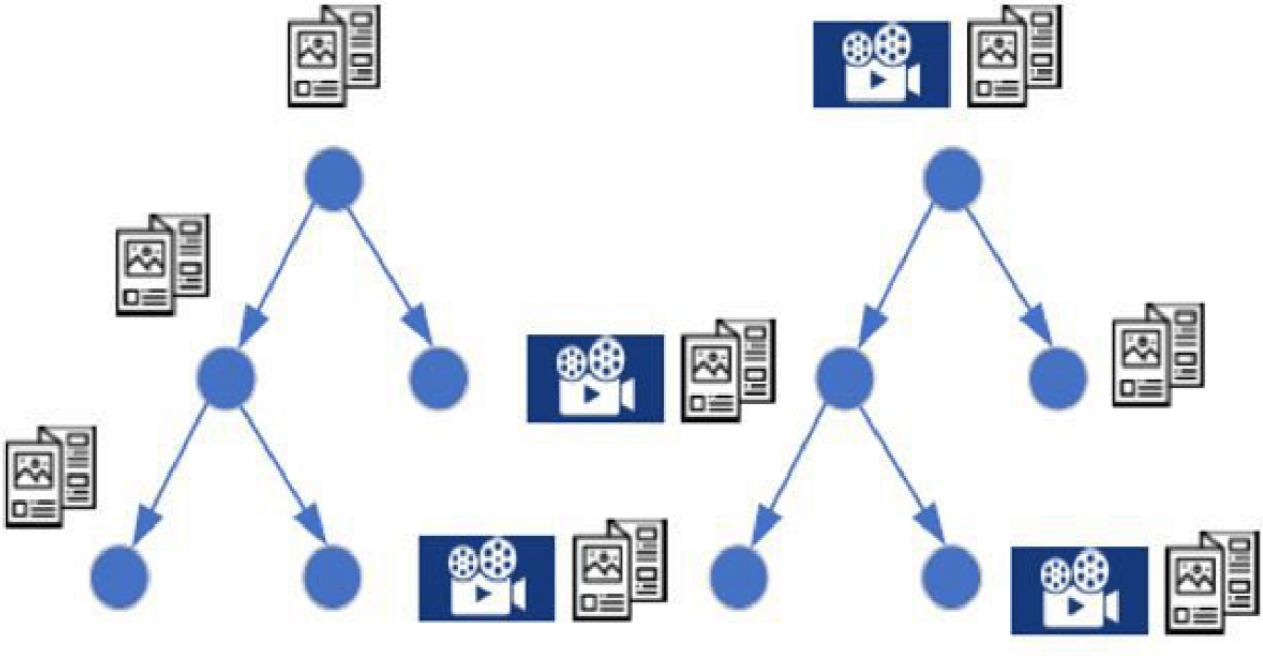
Randomization and Masking

Respondents were asked to complete a web-based questionnaire, in which questions on IDM were asked before and after the control or intervention condition. The control group was asked to read the information brochure regarding the screening program that is currently sent with the screening invitation. The intervention group was asked to read the same brochure and watch the CSEV. This request was displayed on a web page.

By clicking *Next*, they first received the brochure, and subsequently on the next page, the CSEV was displayed.

RDS starts with a convenient, ideally diverse, sample of members of the population called seeds [11]. Seeds were asked to complete a questionnaire and recruit a number of their peers to complete the same questionnaire. The successfully recruited peers were then also asked to recruit a number of peers. This recruitment process was continued until the calculated sample size was reached. Unique tokens were used to follow who recruited whom and draw recruitment trees. Each new respondent was randomly assigned to either the control or the intervention condition (ie, individual-level randomization; Figure 1).

Figure 1. Study design: respondent-driven sampling where each new respondent was randomly assigned to either the control or intervention group.



Study Population and Recruitment

The inclusion criteria for respondents were as follows: the women must be (1) aged 30 to 60 years, (2) born in Turkey or Morocco and have at least one parent born in Turkey or Morocco (first-generation immigrants) or born in the Netherlands and have at least one parent born in Turkey or Morocco (second-generation immigrants), and (3) living in the Netherlands.

Seeds were recruited via several social media platforms, such as (1) public and private women's groups on Facebook, (2) LinkedIn pages of the involved researchers, (3) the foundation called the Association Moroccan Doctors Netherlands, (4) the participating video producer Zouka Media, and (5) Instagram, wherein we contacted several influencers with many Turkish- and Moroccan-Dutch female followers and asked them to share the questionnaire via their story or bio. Throughout the study, we used paper- and web-based flyers and web-based infographics to promote and share the link to the questionnaire. The flyers and infographics were spread among offline community organizations, foundations, and mosques, as well as web-based platforms, such as LinkedIn and Facebook.

After completion of the questionnaire, respondents were asked to invite—through WhatsApp, platforms such as Instagram, and/or SMS text messaging—a maximum of 20 women from their social network to complete the same questionnaire. Via email, reminders were sent to complete and/or forward the questionnaire and to encourage respondents to remind their peers to complete the questionnaire (after 1 week of no participation of at least one peer). To prevent respondents from potentially influencing each other's answers, respondents were explicitly requested not to discuss their answers or watch the CSEV with others. Initially, an incentive of €10 (US $9.69) was awarded to every respondent who completed the questionnaire herself and peer recruited 2 other women who also completed the questionnaire. From March 3, 2021, to further stimulate peer recruitment, an incentive of €15 (US $14.53) was awarded to every respondent who completed the questionnaire herself and peer recruited 1 other woman who also completed the questionnaire.

The Questionnaire

We developed a questionnaire for measuring IDM based on the rational decision model, which supposes that decision-making is based on a proper understanding of the potential benefits and adverse effects of cancer screening (decision-relevant knowledge) in the context of personal situations and preferences (attitude) [12]. The questionnaire contained 52 questions regarding sociodemographic characteristics, previous CC screening participation, knowledge of CC screening, attitude toward CC screening, and intention to participate in the next CC screening round. The questionnaire (in Dutch) can be found in Multimedia Appendix 1. All questions were closed ended, except for the month and year of birth, the 4 digits of the postal code, and the size of their social network on the web. We asked

questions on knowledge, attitude, and intention for clinician-based sampling, whereas for self-sampling, we included questions on awareness, perceptions, and intention. The rationale for this difference was that the self-sampling method was only introduced in 2017, which meant that not every woman was aware of its existence. Therefore, instead of assessing their knowledge and attitude, we questioned their awareness and perceptions of self-sampling. Knowledge of CC screening was measured using 3 questions about the subsequent steps following a test result and the possibility of false-positive test results, with scores ranging from 0 to 4. Attitude toward CC screening was measured using 10 questions, with scores ranging from 0 to 10. These scores were transformed to 0 to 100 scores to facilitate interpretation, following an earlier study by Korfage et al [13]. In agreement with van den Berg et al [14] and Korfage et al [13], we classified scores in the range of 45 to 55 as a neutral attitude. Scores <45 were classified as having a negative attitude, whereas scores >55 were classified as having a positive attitude. Intention was measured by asking the respondents whether they intended to participate in the next CC screening round. All questions regarding attitude and intention had 3 response options: “Yes,” “I do not know,” and “No.”

Following earlier research, we combined knowledge, attitude, and intention to calculate IDM (yes or no) [4,8]. An informed decision was defined as having adequate knowledge (total score≥3.0), either a positive attitude (total score>55.0) and a positive intention or a negative attitude (total score<45.0) and a negative intention. All other combinations were defined as an uninformed decision.

The questionnaire was made available in Dutch, Turkish, and Moroccan-Arabic languages. As first-generation Turkish- and Moroccan-Dutch immigrants have low reading abilities, audio recordings in Dutch, Turkish, Moroccan-Arabic, and Moroccan-Berber (a spoken language) languages were made available. To ensure understandability, the questionnaire was extensively pretested among 4 low-literate Turkish- and Moroccan-Dutch women. It took women approximately 15 minutes to complete the questionnaire.

Culturally Sensitive Educational Videos

We developed 3 CSEVs in collaboration with the video producer and 8 Turkish- and Moroccan-Dutch peer educators and actresses. As all respondents received the brochure containing cognitive information on CC screening, we focused the video on affective information related to CC screening (ie, experiences and fears). Turkish- and Moroccan-Dutch women especially need information on the practical, emotional, cultural, and religious aspects of CC screening [5]. Therefore, the CSEVs emphasized on 3 themes regarding clinician-based sampling and ensured balanced content in terms of possible benefits and adverse effects. The themes included “more assurance regarding health and the ability to prevent treatment, surgery, or death, and because of this, being there for their children”; “according to the Islam, a woman should take good care of her health”; and “anxiety, shame, and privacy.” For self-sampling, 2 themes were included, namely “it is easy and not painful to perform self-sampling” and “trust in themselves to correctly perform self-sampling and trust in the test result.” The CSEV was available in Turkish, Moroccan-Arabic, and Moroccan-Berber (all with Dutch subtitles) languages. Moroccan-Dutch respondents could choose either a Moroccan-Arabic–spoken or Moroccan-Berber–spoken video.

To verify whether the CSEVs were understandable and culturally appropriate, discussions on the web were held among experts on language, communication, culture, and CC (screening). The CSEVs were also pilot-tested in a small sample of Turkish- and Moroccan-Dutch women to verify whether the feasibility, content, and layout matched their needs and requirements. Through automatic registration by the questionnaire software, we measured whether and how long the respondents consulted the brochure (in both the control and intervention groups) and whether the intervention group actually watched the CSEV.

All CSEVs are available on the official webpage of the Dutch National Institute for Public Health and the Environment [15]. Further details regarding the development and tailoring of the CSEVs are reported elsewhere [16].

Sample Size Calculation

We used a 2-sided test and assumed a binomial distribution, 95% CI, 80% power, and an absolute change of 10% in IDM. Therefore, 776 Turkish- and 794 Moroccan-Dutch women (in total; both the control and intervention groups) were needed. This absolute change of 10% in IDM was based on a previously reported study using a developed CSEV and observing an increase of 11% in IDM regarding prenatal screening among pregnant ethnic minority women in the Netherlands [8].

Statistical Analysis

The flow of respondents’ inclusion was visualized. Possible insincere respondents (ie, those that probably participated for incentives only) were excluded from the data and were not eligible for an incentive whenever one of the following criteria was met: (1) the respondent *and* her recruitee completed the questionnaire in <5 minutes or (2) the respondent *or* her recruitee completed the questionnaire in <5 minutes, *and* there was <5 minutes between the start of the 2 participations. Respondents who indicated no migration background, indicated a migration background other than Turkish or Moroccan, or did not indicate their country of birth and/or that of their parent or parents, and those aged <30 or >60 years were also excluded.

Descriptive statistics were used to provide an overview of the sample characteristics and the proportion of respondents who viewed the brochure and CSEV. To analyze the potential additional effect of the CSEV compared with that of the brochure only, we conducted intention-to-treat analyses [17]. We assessed the differences in knowledge (or awareness in the case of self-sampling), attitude (or perceptions in the case of self-sampling), intention, and IDM (only for clinician-based sampling) between the control and intervention groups after the control or intervention condition using chi-square tests or Fisher exact tests.

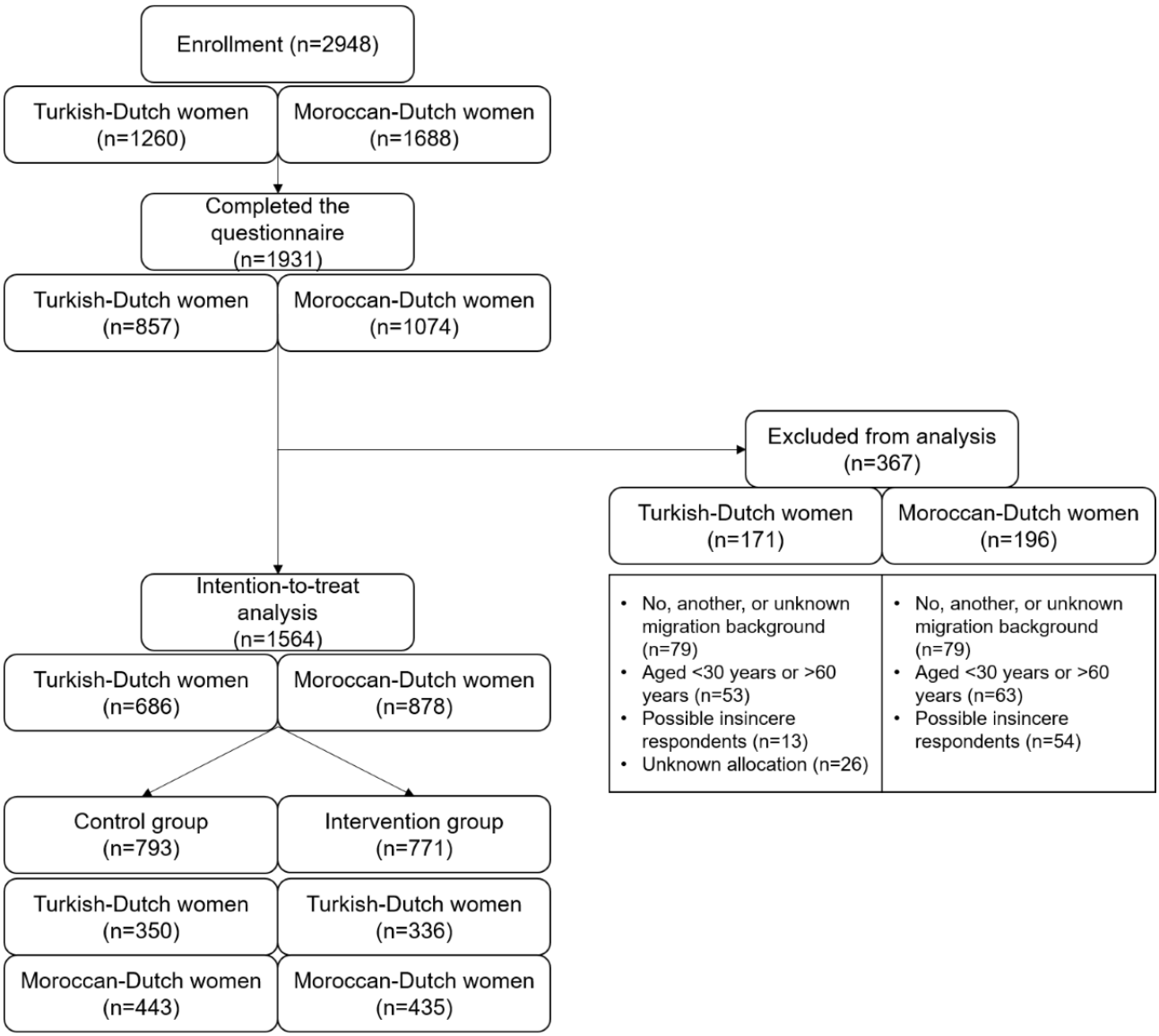
As a post hoc analysis, we explored the open-field comments stated by the respondents at the end of our questionnaire to explain the differences found between Turkish- and

Moroccan-Dutch women. A 2-sided *P* value of <.05 was considered statistically significant. All analyses were performed using the statistical software R (R Foundation for Statistical Computing; version 4.0.2).

Ethics Approval and Consent to Participate

After the Medical Ethics Review Committee of the University Medical Centre Utrecht confirmed that the Medical Research Involving Human Subjects Act does not apply to this study (nr: 20/105), we registered the trial at the International Clinical Trial Registry Platform (trial ID: NL8453). Respondents were informed about the study (but did not know that there was a control group and an intervention group) and were asked to give their digital informed consent.

Figure 2. Flow diagram of the recruitment and response of respondents.



Results

Flow of the Inclusion of Respondents

Of the 2948 respondents that started the questionnaire, 1931 (65.5%) completed it. After excluding 367 (19.01%) respondents, 1564 (80.99%) respondents were included in the analysis, of which 686 (43.86%) respondents were Turkish-Dutch women and 878 respondents (56.14%) were Moroccan-Dutch women: 793 (50.7%) respondents in the control group (350/793, 44.1% Turkish and 443/793, 55.9% Moroccan) and 771 respondents (49.3%) in the intervention group (336/771, 43.6% Turkish and 435/771, 56.4% Moroccan; Figure 2).

Sample Characteristics

The final sample (n=1564) consisted of 686 (43.86%) Turkish-Dutch women and 878 (56.14%) Moroccan-Dutch women (Table 1). Most respondents in both groups were aged between 30 and 39 years and were highly educated (295/686, 43% and 454/878, 51.7%, respectively), and 8% (55/686) and 12% (105/878) of the respondents had no official or primary education, respectively. Overall, 59.9% (411/686) of the Turkish women and 56.4% (495/878) of the Moroccan women were second-generation immigrants. Their social network on the web (ie, other Turkish- or Moroccan-Dutch women aged 30-60 years) was mostly between 11 and 49 women, and 3% of the respondents had no social network on the web.

Table 1. Sample characteristics of the Turkish- and Moroccan-Dutch respondents[a].

Characteristics	Turkish (n=686), n (%)	National data proportions[a] (Turkish), %	Moroccan (n=878), n (%)	National data proportions[a] (Moroccan), %
Age (years)				
30-39	418 (60.9)	36	455 (51.8)	40
40-49	189 (27.6)	37	328 (37.4)	37
50-60	92 (13.4)	27	95 (10.8)	23
Missing value	0 (0)	N/A[b]	0 (0)	N/A
Educational level				
No official education or primary school	82 (12.0)	44[c]	68 (7.7)	43[c]
Secondary school	110 (16.0)	N/A	136 (15.5)	N/A
Vocational education	198 (28.9)	33	219 (24.9)	39
Higher education	295 (43.0)	23	454 (51.7)	18
Missing value	1 (0)	N/A	1 (0)	N/A
Generation				
First	275 (40.1)	72	383 (43.6)	73
Second	411 (59.9)	28	495 (56.4)	27
Missing value	0 (0)	N/A	0 (0)	N/A
Size of their social network on the web				
0	17 (2.5)	N/A	24 (2.7)	N/A
1-10	180 (26.2)	N/A	160 (18.2)	N/A
11-49	273 (39.8)	N/A	412 (46.9)	N/A
50-99	96 (14.0)	N/A	157 (17.9)	N/A
100-249	74 (10.8)	N/A	91 (10.4)	N/A
250-499	33 (4.8)	N/A	29 (3.3)	N/A
≥500	13 (1.9)	N/A	5 (1.0)	N/A
Missing value	0 (0)	N/A	0 (0)	N/A
Language in which the questionnaire was completed				
Dutch	432 (63.0)	N/A	758 (86.3)	N/A
Turkish or Arabic	234 (34.1)	N/A	2 (0)	N/A
Missing value (due to technical failure)	20 (2.9)	N/A	118 (13.4)	N/A
Previous CC[d] screening participation				
Every 5 years	305 (44.5)	N/A	433 (49.3)	N/A
Not every 5 years	106 (15.5)	N/A	125 (14.2)	N/A
Never	275 (40.1)	N/A	320 (36.4)	N/A
Missing value	0 (0)	N/A	0 (0)	N/A

[a]Extracted from databases [18-20].

[b]N/A: not applicable.

[c]Includes no official education, primary school, and secondary school.

[d]CC: cervical cancer.

In total, 40.1% (275/686, Turkish) and 36.4% (320/878, Moroccan) of the respondents indicated that they had never participated in CC screening before, 44.5% (305/686, Turkish) and 49.3% (433/878, Moroccan) of the respondents reported to have participated in CC screening once every 5 years, and 15.5% (106/686, Turkish) and 14.2% (125/878, Moroccan) of the

respondents participated irregularly. The respondents represented a wide geographic area across the Netherlands (Figures S1 and S2 in Multimedia Appendix 1).

Among the Turkish-Dutch women, 33.1% (116/350) of the control respondents and 40.5% (136/336) of the intervention respondents viewed the brochure (not statistically significant). Of the intervention respondents, 96.1% (323/336) of the respondents viewed the CSEV. Among the Moroccan-Dutch women, 28.2% (125/443) of the control respondents and 37.9% (165/435) of the intervention respondents viewed the brochure (P=.003). Of the intervention respondents, 84.4% (367/435) of the respondents viewed the CSEV.

Knowledge of CC Screening

Turkish-Dutch respondents with sufficient knowledge of CC screening increased from 54.6% (191/350) to 68.3% (239/350) in the control group (+13.7% absolute change; P<.001) and from 49.1% (165/336) to 63.7% (214/336) in the intervention group (+14.6%; P<.001). Moroccan-Dutch respondents with sufficient knowledge increased from 61.4% (272/443) to 78.8% (349/443) in the control group (+17.4%; P<.001) and from 65.7% (286/435) to 77.5% (337/435) in the intervention group (+11.8%; P<.001). In terms of knowledge, the CSEV did not show a significant effect above the information brochure for either group (see Tables S1 and S2 in Multimedia Appendix 1).

Attitude Toward CC Screening

Turkish-Dutch respondents with a positive attitude toward CC screening decreased from 70% (245/350) to 67.1% (235/350) in the control group (−2.9%; not statistically significant) and from 66.7% (224/336) to 66.4% (223/336) in the intervention group (−0.3%; not statistically significant). Moroccan-Dutch respondents with a positive attitude increased from 64.6% (286/443) to 68.4% (303/443) in the control group (+3.8%; not statistically significant) and from 65.1% (283/435) to 74.3% (323/435) in the intervention group (+9.2%; P=.004). Overall, there was no added effect of the CSEV on the attitude toward CC screening among Turkish-Dutch women (P=.89; Table S3 in Multimedia Appendix 1). We found that Moroccan-Dutch women in the intervention group more often had a positive attitude toward CC screening compared with the control group, although this difference was not statistically significant (P=.07; Table S4 in Multimedia Appendix 1). Moroccan-Dutch women in the intervention group who had never participated in CC screening before had significantly more often a positive attitude toward CC screening compared with the control group (P=.01).

Intention and IDM Regarding CC Screening Participation

Both the control and intervention groups had more often a positive intention after consulting the brochure or the brochure and CSEV in both Turkish- and Moroccan-Dutch women (Table 2). An increase was observed among Turkish-Dutch women from 78.3% (274/350) to 82.6% (289/350) in control respondents (+4.3%; not statistically significant) and from 79.2% (266/336) to 84.5% (284/336) in intervention respondents (+5.3%; not statistically significant). The same holds true for Moroccan-Dutch women: from 79.9% (354/443) to 86% (381/443) in control respondents (+6.1%; P=.02) and from 80% (348/435) to 86.9% (378/435) in intervention respondents (+6.9%; P=.008). However, the CSEV did not have a statistically significant added effect above the brochure in terms of intention.

In general, women made more often an informed decision after the control or intervention condition among Turkish- and Moroccan-Dutch women (Table 2). Of the control respondents, IDM increased from 38.6% (135/350) to 44.3% (155/350) in Turkish-Dutch women (+5.7%; not statistically significant) and from 43.8% (194/443) to 53.7% (238/443) in Moroccan-Dutch women (+9.9%; P=.004). The same holds true for intervention respondents; we saw an increase in IDM from 34.5% (116/336) to 42.9% (144/336) in Turkish-Dutch women (+8.4%; P=.03) and from 44.6% (194/435) to 58.9% (256/435) in Moroccan-Dutch women (+14.3%; P<.001). However, the CSEV did not have a statistically significant added effect above the brochure in terms of IDM (Tables 2 and 3).

Table 2. Intention and informed decision-making (IDM) regarding cervical cancer (CC) screening participation in the control and intervention groups, before and after reading the brochure (control) or reading the brochure and watching the culturally sensitive educational video (intervention).

Characteristics	Population					
	Turkish-Dutch women			Moroccan-Dutch women		
	Control group (n=350), n (%)	Intervention group (n=336), n (%)	*P* value	Control group (n=443), n (%)	Intervention group (n=435), n (%)	*P* value
Intention to participate in CC screening (before)						
Positive	274 (78.3)	266 (79.2)	.85	354 (79.9)	348 (80.0)	>.99
Neutral	60 (17.1)	58 (17.3)	>.99	65 (14.7)	68 (15.6)	.76
Negative	16 (4.6)	12 (3.6)	.64	24 (5.4)	19 (4.4)	.57
Intention to participate in CC screening (after)						
Positive	289 (82.6)	284 (84.5)	.56	381 (86.0)	378 (86.9)	.77
Neutral	48 (13.7)	41 (12.2)	.63	45 (10.2)	41 (9.4)	.80
Negative	13 (3.7)	11 (3.3)	.92	17 (3.8)	16 (3.7)	>.99
IDM (before)						
Yes	135 (38.6)	116 (34.5)	.31	194 (43.8)	194 (44.6)	.86
No	215 (61.4)	220 (65.5)	N/A[a]	249 (56.2)	241 (55.4)	N/A
IDM (after)						
Yes	155 (44.3)	144 (42.9)	.76	238 (53.7)	256 (58.9)	.14
No	195 (55.7)	192 (57.1)	N/A	205 (46.3)	179 (41.1)	N/A

[a]N/A: not applicable.

Table 3. Informed decision-making regarding cervical cancer (CC) screening participation in the control and intervention groups, after reading the brochure (control) or reading the brochure and watching the culturally sensitive educational video (intervention).

Characteristics	Population					
	Turkish-Dutch women			Moroccan-Dutch women		
	Control group (N=350), n (%); uninformed[a]	Intervention group (N=336), n (%); uninformed[a]	*P* value	Control group (N=443), n (%); uninformed[a]	Intervention group (N=435), n (%); uninformed[a]	*P* value
Age (years)						
30-39	124 (35.4)	112 (33.3)	.62	107 (24.2)	89 (20.5)	.22
40-49	47 (13.4)	52 (15.5)	.51	72 (16.3)	62 (14.3)	.47
50-60	24 (6.9)	28 (8.3)	.56	26 (5.9)	28 (6.4)	.83
Educational level						
No official education or primary school	23 (6.6)	32 (9.5)	.20	18 (4.1)	25 (5.7)	.32
Secondary school	29 (8.3)	39 (11.6)	.18	34 (7.7)	25 (5.7)	.31
Vocational education	62 (17.7)	54 (16.1)	.64	54 (12.2)	44 (10.1)	.39
Higher education	80 (22.9)	67 (19.9)	.40	99 (22.3)	85 (19.5)	.35
Generation						
First	74 (21.1)	89 (26.5)	.12	96 (21.7)	87 (20.0)	.60
Second	121 (34.6)	103 (30.7)	.31	109 (24.6)	92 (21.1)	.26
Previous CC screening participation						
Every 5 years	58 (16.6)	71 (21.1)	.15	70 (15.8)	56 (12.9)	.25
Not every 5 years	37 (10.6)	31 (9.2)	.64	26 (5.9)	31 (7.1)	.54
Never	100 (28.6)	90 (26.8)	.66	109 (24.6)	92 (21.1)	.26

[a]Uninformed: The number of women classified as being uninformed.

Self-sampling

No statistically significant differences were found in awareness, perceptions, and intention regarding self-sampling when comparing the control and intervention groups among Turkish-Dutch women (Table S9 in Multimedia Appendix 1).

More Moroccan-Dutch respondents thought that self-sampling was easy to perform in the intervention group than in the control group (284/435, 65.3% vs 252/443, 56.9%; *P*=.04). In addition, fewer respondents in the intervention group thought that self-sampling would be painful compared with the control group (59/435, 13.6% vs 82/443, 18.5%; *P*=.05; Table S10 in Multimedia Appendix 1).

Discussion

Principal Findings

This study evaluated the effect of a CSEV on knowledge, attitude, intention, and IDM regarding CC screening among Turkish- and Moroccan-Dutch women aged 30 to 60 years. The CSEV was watched far more often than the brochure was read when both were offered together, and the intervention group who watched the video also studied the brochure more often than the control group did. The brochure had a significant positive influence on IDM, whereas the CSEV had an added effect on the attitude toward CC screening, especially in Moroccan-Dutch women. These women more often had a positive attitude toward CC screening compared with the control group who had read only the brochure. This was especially the case among women who had never participated in CC screening before. On the basis of the open-field comments of Turkish-Dutch respondents, we think we can explain why this effect was not visible in this group. It appeared that some of the Turkish-Dutch respondents were offended by the fact that in the Turkish video, the actress who played having a negative screening attitude was wearing a headscarf.

Comparison With Prior Work

In line with our results in the control group, a previous study among Dutch women invited for breast cancer screening also found that reading the brochure enhanced IDM [21]. Earlier randomized controlled trials that strived to enhance IDM regarding cancer screening often developed a decision aid, in which information was presented differently compared with the standard letter or brochure [22-26]. These studies tended to target knowledge instead of the attitudes we aimed at. In line with our study, an earlier randomized controlled trial in Germany among all targeted women without a Turkish migration background also compared the standard information brochure for breast cancer screening with a newly developed decision aid [27]. In contrast to our study, more respondents in the intervention group were knowledgeable compared with those in the control group. This seems to be related to the fact that

the same information was presented in both the groups, but only visually, instead of textually, in the intervention group versus the control group. We did not include any factual medical information in the CSEV and did not target women's knowledge. In the United Kingdom, a similar intervention study regarding participation in lung cancer screening among smokers also used a video and found that it improved knowledge and reduced decisional conflict [28]. However, this video was also targeted at increasing knowledge instead of improving screening attitudes.

Implications for Practice and Policy

We recommend developing videos that incorporate information provided in the current brochure, as many Turkish- and Moroccan-Dutch women do not read the brochure (thoroughly) or are simply unable to read it [5]. In line with this study, a video has been shown to be more engaging and attractive than textual information [29]. Considering that approximately one-third of the control group consulted the brochure, the effect of the brochure on IDM might be greater if the brochure was studied more often and in more detail. We expect that in the context of this study, respondents were more likely to read the brochure (intensively) than those who received it with the invitation (ie, the Hawthorne effect). Therefore, we recommend presenting the CSEV to all women through the invitation letter, for example, using a weblink or a QR code, so that the CSEV and all other web-based materials can be accessed easily. We propose to consider using the CSEV in mosques, community centers, and educational meetings regarding (women's) health for women with limited digital skills. Other options include distributing the CSEV in women's groups on Facebook or broadcasting the CSEV on a loop in the waiting room at the GP's office.

Women are invited to undergo CC screening every 5 years and might not be interested to search for or gather information every time they are invited. Therefore, in addition to evaluating different modes of delivering visual information, we recommend that research be performed on the use of different distribution channels to reach uninformed women, such as social media and involvement of influencers, key figures, informants, and close-knit community groups that were used in this study.

In October 2021, the Dutch Health Council recommended offering self-sampling as an equivalent alternative to clinician-based sampling and sending the self-sampling kit together with the invitation [30]. Owing to the CSEV, more Moroccan-Dutch respondents thought that self-sampling was easy to perform and fewer respondents thought that self-sampling would be painful. Therefore, sending the self-sampling kit with the invitation should concur with implementing our CSEV. Overall, as a short intervention that is easily implemented, our CSEV represents an efficient way to enhance screening attitudes and facilitate IDM among immigrant women.

Strengths and Limitations

One major strength of this study was its design as a randomized intervention study. Worldwide, this study also had one of the largest samples successfully recruited using web-based RDS [31]. In addition, our CSEVs were systematically developed based on extensive qualitative and quantitative research among Turkish- and Moroccan-Dutch women [5]. The brochure that we used in our study was sent to all women aged 30 to 60 years by the regional screening organizations. This brochure has been used in practice since November 2016 and is considered "usual care," and it openly discusses potential benefits and harms of CC screening. Therefore, we deliberately used the CSEV as an addition to the brochure to facilitate one's individual thinking process and/or discussion with other women and not as a replacement intervention. Our CSEV can now be easily added to the existing invitation materials. More importantly, our CSEV includes other more affective aspects, which are not incorporated in the brochure but are needed for the Turkish- and Moroccan-Dutch women to be able to make a conscious decision on their CC screening participation [5].

However, a number of limitations should also be addressed. First, owing to the web-based delivery of the questionnaire, we sampled a greater number of women who were aged 30 to 39 years, were second-generation immigrants, and were highly educated Turkish- and Moroccan-Dutch women compared with the national data set of 2020 of Statistics Netherlands [18-20]. However, the 2 randomized groups were comparable, and 12% (82/686) and 8% (70/878) of the Turkish- and Moroccan-Dutch respondents reported no official education or completed primary school, respectively. In addition, regarding previous CC screening participation, we did find similar rates of at least one participation in CC screening of 60% (412/686) and 64% (562/878) of the respondents in Turkish- and Moroccan-Dutch women versus 64% and 53% of the respondents, respectively, in previous reports [6].

Second, the time elapsed between the previous screening invitation and the questionnaire administration, which varied largely among our respondents, might have affected the experienced relevance of the decision-making questions and the previously existing knowledge. However, this heterogeneity is likely to play a similar role (if it does at all) in both the control and intervention groups because of the randomization performed.

Third, the women who participated in our study might have been different from those who did not participate in the study. For example, they could be more interested in CC screening as a topic and be more informed about it than nonparticipating women. Nevertheless, as we used incentives for successful peer recruitment, this might also have been the reason that some respondents participated in the study rather than being interested in CC screening. In addition, this possible selection bias is likely to be present in both the control and intervention groups and should not affect the evaluation of the CSEV.

Fourth, our knowledge construct contained only some facts about CC screening (ie, the process after a negative or positive test result and the possibility of false-positive test results). Although these have been carefully selected, they do not cover the entire spectrum of decision-relevant information (eg, hrHPV as the causative agent of CC and its transmission route) and can only indicate some deficits. Because of the use of RDS, and thus requesting women to successfully recruit others, we aimed

to burden the respondents as less as possible and, therefore, kept the questionnaire as short as possible.

Fifth, health literacy (ie, the degree to which individuals have the ability to find, understand, and use information and services to take informed health-related decisions and actions for themselves and others) is crucial to make informed health-related decisions. The immigrants are less capable of applying IDM, as they have lower health literacy levels compared with nonimmigrants [32]. It would have been interesting to assess health literacy levels of individuals to compare the effect of our CSEV among those with limited and adequate health literacy levels.

Sixth, to further explore the differences found between Turkish- and Moroccan-Dutch women and their attitudes, thoughts, and views regarding the current information brochure and the CSEV, it would have been highly relevant to conduct follow-up interviews or focus groups, shortly after the end of our randomized intervention study.

Finally, we based the content of the CSEVs on our earlier conducted focus groups among offline-recruited Turkish- and Moroccan-Dutch women [5]. Because of the measures taken for the COVID-19 pandemic (eg, nationwide lockdowns), we were unable to approach potential respondents face-to-face and recruit them offline. The respondents were also unable to recruit peers offline unless they were household members. This resulted in a web-based–only, relatively young, mostly second-generation sample of Turkish- and Moroccan-Dutch women. It would be highly relevant to evaluate the CSEVs in an offline setting, which is comparable with our previous study [5]. We believe that CSEVs could affect IDM (greater) in such a setting for which the CSEVs were tailored during the development process.

Conclusions

This randomized intervention study has demonstrated that a CSEV positively affected CC screening attitudes, especially among Moroccan-Dutch women. Women who were offered both the brochure and CSEV consulted the brochure more often than those who received the brochure only. The CSEV was also watched far more often than the brochure was read. Therefore, the CSEV can be widely distributed through offline and web-based channels, in addition to the current information materials.

Acknowledgments

The authors express their gratitude to all Turkish- and Moroccan-Dutch women who participated in the study and those who worked with them to design, develop, and pretest the questionnaire and the culturally sensitive educational video (CSEV). The authors also thank Abdelkarim and Asma El-Fassi (video producer and director) for designing and developing our CSEVs and their help in recruiting Turkish- and Moroccan-Dutch women. The authors would like to thank their own network, influencers, key figures and informants, Facebook groups, mosques, community centers, and foundations that helped with the recruitment of Turkish- and Moroccan-Dutch women. The authors are thankful to the Association Moroccan Doctors Netherlands and Foundation Health Immigrants Netherlands for their help during the development of the questionnaire and the CSEVs and recruitment of Turkish- and Moroccan-Dutch women. The use of the web-based respondent-driven sampling software system developed by the Karolinska Institutet, UMC Utrecht, and the Dutch National Institute for Public Health and the Environment is gratefully acknowledged. The authors thank Elle Langens of the Centre for Population Screening of the Dutch National Institute for Public Health and the Environment for her valuable input during the development of the questionnaire and the CSEVs. This study was part of a larger research project called *the FEMININE study*, which was funded by The Netherlands Organization for Health Research and Development (nr: 531002030). The funder had no role in the study design, data collection, data analysis, data interpretation, writing of the report, and the decision to submit the article for publication.

Authors' Contributions

AT, MvdM, MLS, JvS, and NH conceptualized the study and acquired funding. AK, MNÇ, and NH recruited the respondents and collected the data. NH performed the formal analysis. NH wrote and prepared the first draft of the manuscript. MLS, JvS, RC, MB, AK, MNÇ, AT, and MvdM reviewed and revised the manuscript. All the authors read and approved the final version of the manuscript. All the authors had full access to all the data in the study and are responsible for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 397 KB - jmir_v24i10e35962_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH Checklist (V 1.6.1).
[DOCX File , 804 KB - jmir_v24i10e35962_app2.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Yang DX, Soulos PR, Davis B, Gross CP, Yu JB. Impact of widespread cervical cancer screening: number of cancers prevented and changes in race-specific incidence. Am J Clin Oncol 2018 Mar;41(3):289-294 [FREE Full text] [doi: 10.1097/COC.0000000000000264] [Medline: 26808257]
3. Arbyn M, Ronco G, Anttila A, Meijer CJ, Poljak M, Ogilvie G, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 2012 Nov 20;30 Suppl 5:F88-F99. [doi: 10.1016/j.vaccine.2012.06.095] [Medline: 23199969]
4. Marteau TM, Dormandy E, Michie S. A measure of informed choice. Health Expect 2001 Jun;4(2):99-108 [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00140.x] [Medline: 11359540]
5. Hamdiui N, Marchena E, Stein ML, van Steenbergen JE, Crutzen R, van Keulen HM, et al. Decision-making, barriers, and facilitators regarding cervical cancer screening participation among Turkish and Moroccan women in the Netherlands: a focus group study. Ethn Health 2022 Jul;27(5):1147-1165. [doi: 10.1080/13557858.2020.1863921] [Medline: 33412893]
6. Steens A, Wielders CC, Bogaards JA, Boshuizen HC, de Greeff SC, de Melker HE. Association between human papillomavirus vaccine uptake and cervical cancer screening in the Netherlands: implications for future impact on prevention. Int J Cancer 2013 Mar 15;132(4):932-943 [FREE Full text] [doi: 10.1002/ijc.27671] [Medline: 22689326]
7. Integratie; migratieachtergrond, 2006-2011. CBS. 2011. URL: https://opendata.cbs.nl/#/CBS/nl/dataset/80270ned/table?ts=1598528180747 [accessed 2021-12-24]
8. Peters IA, Posthumus AG, Reijerink-Verheij JC, Van Agt HM, Knapen MF, Denktaş S. Effect of culturally competent educational films about prenatal screening on informed decision making of pregnant women in the Netherlands. Patient Educ Couns 2017 Apr;100(4):776-782. [doi: 10.1016/j.pec.2016.11.007] [Medline: 27887753]
9. Hamdiui N, van Steenbergen J, Rocha LE, Meiberg A, Urbanus A, Hammou NA, et al. Hepatitis B screening among immigrants: how to successfully reach the Moroccan community. J Viral Hepat 2021 Dec;28(12):1759-1762 [FREE Full text] [doi: 10.1111/jvh.13604] [Medline: 34467605]
10. Odé AW. Ethnic-cultural and Socio-economic Integration in the Netherlands: A Comparative Study of Mediterranean and Caribbean Minority Groups. Assen, The Netherlands: Uitgeverij Van Gorcum; 2002.
11. Tyldum G, Johnston LG. Applying Respondent Driven Sampling to Migrant Populations: Lessons from the Field. Amsterdam, The Netherlands: Springer; 2014.
12. Simon HA. A behavioral model of rational choice. Q J Econ 1955 Feb;69(1):99-118. [doi: 10.2307/1884852]
13. Korfage IJ, van Ballegooijen M, Wauben B, Habbema JD, Essink-Bot ML. Informed choice on Pap smear still limited by lack of knowledge on the meaning of false-positive or false-negative test results. Patient Educ Couns 2011 Nov;85(2):214-218 [FREE Full text] [doi: 10.1016/j.pec.2010.12.014] [Medline: 21269793]
14. van den Berg M, Timmermans DR, Ten Kate LP, van Vugt JM, van der Wal G. Are pregnant women making informed choices about prenatal screening? Genet Med 2005;7(5):332-338 [FREE Full text] [doi: 10.1097/01.gim.0000162876.65555.ab] [Medline: 15915085]
15. Video's over het bevolkingsonderzoek. Rijksinstituut voor Volksgezondheid en Milieu. 2021. URL: https://www.rivm.nl/bevolkingsonderzoek-baarmoederhalskanker/videos-marokkaans-turks [accessed 2021-12-24]
16. Hamdiui N, Bouman MP, Stein ML, Crutzen R, Keskin D, Afrian A, et al. The development of a culturally sensitive educational video: How to facilitate informed decisions on cervical cancer screening among Turkish- and Moroccan-Dutch women. Health Expect (forthcoming) 2022 Jul 21 [FREE Full text] [doi: 10.1111/hex.13545] [Medline: 35866200]
17. van Buuren S. Flexible Imputation of Missing Data. 2nd edition. Boca Raton, FL, USA: CRC Press; 2018.
18. Bevolking; leeftijd, migratieachtergrond, geslacht, regio, 1 Jan. 1996-2020. CBS. URL: https://opendata.cbs.nl/#/CBS/nl/dataset/37713/table?ts=1594720128276 [accessed 2020-07-14]
19. Bevolking; geslacht, leeftijd, generatie en migratieachtergrond, 1 Januari. CBS. 2020. URL: https://opendata.cbs.nl/#/CBS/nl/dataset/37325/table [accessed 2020-06-11]
20. Bevolking; onderwijsniveau en migratieachtergrond 2003-2021. CBS. 2020. URL: https://opendata.cbs.nl/#/CBS/nl/dataset/82275NED/table?ts=1594718811823 [accessed 2020-07-14]
21. Kregting LM, van Ravesteyn NT, Spijker W, Dierks T, Aitken CA, Geuzinge HA, et al. Effects of a leaflet on breast cancer screening knowledge, explicit attitudes, and implicit associations. Patient Educ Couns 2020 Jul 02;103(12):2499-2507 [FREE Full text] [doi: 10.1016/j.pec.2020.06.032] [Medline: 32732022]
22. Carter-Harris L, Comer RS, Slaven Ii JE, Monahan PO, Vode E, Hanna NH, et al. Computer-tailored decision support tool for lung cancer screening: community-based pilot randomized controlled trial. J Med Internet Res 2020 Nov 03;22(11):e17050 [FREE Full text] [doi: 10.2196/17050] [Medline: 33141096]
23. Gabel P, Edwards A, Kirkegaard P, Larsen MB, Andersen B. The LEAD trial-the effectiveness of a decision aid on decision making among citizens with lower educational attainment who have not participated in FIT-based colorectal cancer screening in Denmark: a randomised controlled trial. Patient Educ Couns 2020 Feb;103(2):359-368 [FREE Full text] [doi: 10.1016/j.pec.2019.08.029] [Medline: 31451360]

24. Gabel P, Larsen MB, Edwards A, Kirkegaard P, Andersen B. Effectiveness of a decision aid for colorectal cancer screening on components of informed choice according to educational attainment: a randomised controlled trial. PLoS One 2020 Nov 10;15(11):e0241703 [FREE Full text] [doi: 10.1371/journal.pone.0241703] [Medline: 33170877]
25. Pérez-Lacasta MJ, Martínez-Alonso M, Garcia M, Sala M, Perestelo-Pérez L, Vidal C, with the InforMa Group. Effect of information about the benefits and harms of mammography on women's decision making: the InforMa randomised controlled trial. PLoS One 2019 Mar 26;14(3):e0214057 [FREE Full text] [doi: 10.1371/journal.pone.0214057] [Medline: 30913217]
26. Volk RJ, Lowenstein LM, Leal VB, Escoto KH, Cantor SB, Munden RF, et al. Effect of a patient decision aid on lung cancer screening decision-making by persons who smoke: a randomized clinical trial. JAMA Netw Open 2020 Jan 03;3(1):e1920362 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20362] [Medline: 32003822]
27. Reder M, Kolip P. Does a decision aid improve informed choice in mammography screening? Results from a randomised controlled trial. PLoS One 2017 Dec 13;12(12):e0189148 [FREE Full text] [doi: 10.1371/journal.pone.0189148] [Medline: 29236722]
28. Ruparel M, Quaife SL, Ghimire B, Dickson JL, Bhowmik A, Navani N, et al. Impact of a lung cancer screening information film on informed decision-making: a randomized trial. Ann Am Thorac Soc 2019 Jun;16(6):744-751 [FREE Full text] [doi: 10.1513/AnnalsATS.201811-841OC] [Medline: 31082267]
29. Yadav A, Phillips MM, Lundeberg MA, Koehler MJ, Hilden K, Dirkin KH. If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. J Comput High Educ 2011 Feb 11;23(1):15-37. [doi: 10.1007/s12528-011-9042-y]
30. Options for improving population screening for cervical cancer. Health Council of the Netherlands. The Hague, The Netherlands: Health Council of the Netherlands; 2021. URL: https://www.healthcouncil.nl/documents/advisory-reports/2021/10/19/options-for-improving-population-screening-for-cervical-cancer [accessed 2021-12-24]
31. Helms YB, Hamdiui N, Kretzschmar ME, Rocha LE, van Steenbergen JE, Bengtsson L, et al. Applications and recruitment performance of web-based respondent-driven sampling: scoping review. J Med Internet Res 2021 Jan 15;23(1):e17564 [FREE Full text] [doi: 10.2196/17564] [Medline: 33448935]
32. Fransen MP, Harris VC, Essink-Bot ML. Beperkte gezondheidsvaardigheden bij patienten van allochtone herkomst: alleen een tolk inzetten is meestal niet genoeg. Nederlands Tijdschrift Voor Geneeskunde. 2013 Apr 3. URL: https://www.ntvg.nl/artikelen/beperkte-gezondheidsvaardigheden-bij-patienten-van-allochtone-herkomst#:~:text=Wij%20willen%20artsen%20meer%20inzicht,een%20taalbarri%C3%A8re%20niet%20genoeg%20is [accessed 2021-12-24]

Abbreviations

CC: cervical cancer
CONSORT: Consolidated Standards of Reporting Trials
CSEV: culturally sensitive educational video
GP: general practitioner
hrHPV: high-risk human papillomavirus
IDM: informed decision-making
RDS: respondent-driven sampling

Edited by A Mavragani; submitted 24.12.21; peer-reviewed by A Murat Ringot, J Polo; comments to author 13.05.22; revised version received 18.05.22; accepted 19.05.22; published 26.10.22.

Please cite as:
Hamdiui N, Stein ML, van Steenbergen J, Crutzen R, Bouman M, Khan A, Çetin MN, Timen A, van den Muijsenbergh M
Evaluation of a Web-Based Culturally Sensitive Educational Video to Facilitate Informed Cervical Cancer Screening Decisions Among Turkish- and Moroccan-Dutch Women Aged 30 to 60 Years: Randomized Intervention Study
J Med Internet Res 2022;24(10):e35962
URL: https://www.jmir.org/2022/10/e35962
doi:10.2196/35962
PMID:36287585

Original Paper

The Effect of an App-Based Home Exercise Program on Self-reported Pain Intensity in Unspecific and Degenerative Back Pain: Pragmatic Open-label Randomized Controlled Trial

Hannes Weise[1,2,3], MD, DMD; Benedikt Zenner[4], MSc; Bettina Schmiedchen[5], PhD; Leo Benning[5], MPH, MD; Michael Bulitta[6], Dipl Stat; Daniel Schmitz[6], MSc; Kuno Weise[1,2,7], MD, PhD

[1]Institute for Occupational Medicine, Social Medicine and Health Services Research, University Hospital Tübingen, Eberhard-Karls-University Tübingen, Tübingen, Germany

[2]Medical Assessment Institute Tübingen, Tübingen, Germany

[3]Faculty of Medicine, University Hospital Tübingen, Eberhard-Karls-University Tübingen, Tübingen, Germany

[4]Institute of Health Care and Public Management, Hohenheim University, Stuttgart, Germany

[5]Vivira Health Lab, Berlin, Germany

[6]CRM Biometrics, Rheinbach, Germany

[7]Faculty of Medicine, BG-Hospital Trauma Center Tübingen, Eberhard-Karls-University Tübingen, Tübingen, Germany

Corresponding Author:
Hannes Weise, MD, DMD
Institute for Occupational Medicine, Social Medicine and Health Services Research
University Hospital Tübingen
Eberhard-Karls-University Tübingen
Hoppe-Seyler-Straße 3
Tübingen, 72076
Germany
Phone: 49 1727368909
Fax: 49 7071 29 25277
Email: hannes.weise@med.uni-tuebingen.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2023/1/e46512

Abstract

Background: The recommended first-line treatment for unspecific and degenerative back pain consists of movement exercises and patient education.

Objective: Using a pragmatic, randomized controlled trial, we evaluated the effectiveness of a digital home exercise program on self-reported pain intensity compared with the standard of care for physiotherapy.

Methods: Participant recruitment was based on newspaper advertisements and a consecutive on-site assessment for eligibility and enrollment. Participants with unspecific and degenerative back pain aged ≥18 years were randomly assigned in a 1:1 ratio to receive a 12-week stand-alone digital home exercise program or physiotherapy. The digital home exercise program included 4 exercises daily, while physiotherapy included 6 to 12 sessions, depending on the severity of symptoms. The primary outcome was pain, which was assessed using a verbal numerical rating scale. The clinical relevance of pain reduction was assessed using the following thresholds: improvement of at least 1.4 points on the verbal numerical rating scale and a pain reduction of at least 30%.

Results: During the study period, 108 participants were assigned to the intervention group and 105 participants to the control group. The mean difference in pain scores between the 2 groups at 12 weeks was −2.44 (95% CI −2.92 to −1.95; $P<.01$) in favor of the intervention group. The group receiving the digital therapeutic achieved a clinically relevant reduction in pain over the course of the study (baseline vs 12 weeks), with a mean change of −3.35 (SD 2.05) score points or −53.1% (SD 29.5). By contrast,

this change did not reach clinical relevance in the control group (mean −0.91, SD 1.5; −14.6%, SD 25.3). Retention rates of 89.9% in the intervention group and 97.3% in the control group were maintained throughout the study.

Conclusions: The use of the app-based home exercise program led to a significant and clinically relevant reduction in pain intensity throughout the 12-week duration of the program. The intervention studied showed superior improvement in self-reported pain intensity when compared with the standard of care. Given the great demand for standard physiotherapy for unspecific and degenerative back pain, digital therapeutics are evolving into a suitable therapeutic option that can overcome the limitations of access and availability of conventional modes of health care delivery into this spectrum of indications. However, further independent evaluations are required to support the growing body of evidence on the effectiveness of digital therapeutics in real-world care settings.

Trial Registration: German Clinical Trials Register DRKS00022781; https://tinyurl.com/hpdraa89

(J Med Internet Res 2022;24(10):e41899) doi:10.2196/41899

KEYWORDS

back pain; musculoskeletal health; primary care; exercise therapy; digital health; mobile health; mHealth; digital therapeutic; mobile phone

Introduction

Musculoskeletal conditions are among the top drivers of the burden of disease worldwide. In the most recent Global Burden of Disease Study, lower unspecific back pain accounted for 2.5% of all disability-adjusted life years [1]. Although the spectrum of musculoskeletal conditions shows a high prevalence among older individuals, it also accounts for significant direct and indirect health care expenses in other age strata [2]. Hence, health care systems face the challenge of providing adequate and timely care for these conditions. The need for adequate and comprehensive care settings has long been identified [3,4], but the availability of and access to adequate care often remains limited. For the spectrum of unspecific musculoskeletal conditions, physiotherapy and other forms of exercise-based therapies have been described as first-line treatments in international guidelines [5-7]. However, these therapies are often not sufficiently available owing to regulations in health care policy [8], limited availability of and access to care [9,10], as well as challenges regarding the delivery of care [11,12].

In this context, new and innovative approaches are required to develop and sustain a responsive and accessible health care delivery infrastructure. While numerous attempts have been made to digitize components of health care related to musculoskeletal conditions, most have failed to be integrated into existing health care systems and established care delivery pathways [13]. However, after the introduction of the Digital Health Care Act (Digitale-Versorgung-Gesetz) in Germany in 2019, digital health apps, referred to as Digitale Gesundheitsanwendungen (DiGA), were established as a new category of digital therapeutics. These digital therapeutics could receive full market approval from the Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), the German body that assesses pharmaceutical products and medical devices regarding their safety and effectiveness and grants market approval. Since then, approved DiGA have become part of the collectively funded health insurance system and can be prescribed by all licensed physicians and other health care professionals in Germany.

In view of this, ViViRA (ViViRA Health Lab GmbH) is the first self-guided home exercise program for the treatment of degenerative and unspecific back pain that has been approved by the BfArM for use in the collectively funded statutory health insurance system. Thus, ViViRA can be integrated into routine medical care in Germany. For approval as a DiGA by the regulatory bodies, a randomized controlled trial demonstrating effectiveness was conducted. Hence, this publication presents data from a pragmatic, open-label randomized controlled trial that aimed to assess the comparative effectiveness of BfArM-approved DiGA ViViRA against the established standard of care for physiotherapy.

Methods

Trial Design

We conducted a pragmatic randomized controlled trial with 1 intervention group and 1 control group. As the mode of administration of the experimental therapy (ie, in the intervention group) differed significantly from that of the control therapy (ie, in the control group), an open-label design was chosen. The intervention group used the digital therapeutic ViViRA, while the control group received the standard treatment of physiotherapy. Intervention and control therapies were administered in parallel. All study-related data (ie, baseline assessment, primary end point data, and supplementary data) were collected between August 2020 and April 2021. No modifications were made to the trial design after its commencement.

Study Population

Inclusion criteria and requirements for participation in the study were defined as follows: (1) age >18 years; (2) diagnosis of a unspecific or degenerative pain of the lower back (International Classification of Diseases, 10th edition M42.0, M42.1, M42.9, M53.2, M53.8, M53.9, M54.4, M54.5, M54.6, M54.8, M54.9, M99.02, M99.03, M99.04, M99.82, M99.83, M99.84, M99.92, M99.93, and M99.94); (3) a pain score of ≥4 out of 10 based on the verbal numerical rating scale (VNRS) at the time of enrollment, which corresponds to at least moderate pain and is a plausible indicator of therapeutic need in a real-world setting; (4) possession of a mobile device (ie, smartphone or tablet) and

the ability to use such a device; and (5) ability to provide informed consent. The exclusion criteria are outlined in Textbox 1.

Textbox 1. Exclusion criteria to participate in the study.

General

- No pain, pain score ≤3
- Previous movement therapy with a digital therapeutic for musculoskeletal pain
- Use of analgesics before inclusion
- Pregnancy
- Limited legal or insufficient language capacity
- Patients who are not able to follow the exercise protocol; for example, significantly impaired vision or blindness

Internal

- Severe organ failure
  - Condition after heart attack
  - Need for dialysis
  - Cardiovascular decompensation
  - Pulmonary insufficiency
- Inflammation
  - Past or present rheumatological disease
  - Acute inflammatory diseases
  - Feverish condition
- Coagulopathy
  - Thrombosis
  - Blood coagulation disorders including anticoagulant therapy

Musculoskeletal

- Any bone disease
- Injuries or surgery
  - Fresh bone or joint fractures
  - Injury to spinal column, knee, or hip joint
  - Condition after
  - Spine, hip, or joint surgery
  - Osteotomy (an operation to correct the axis of the leg)
  - Arthrodesis (joint stiffening) in 1 of the 2 knee or hip joints
- Inflammatory disease
  - Spinal column or joint inflammatory disease
  - Situation after spinal column or joint inflammatory disease
- Spinal tumor
- Osteochondrosis dissecans
- Bone necrosis
- Hip dysplasia
- Acute instability of the knee or hip joint
- Free joint bodies
- Disc pathology
  - Slipped disc

- Acute herniated disc or other disorder with radiation to the legs (radiculopathy or sensorimotor failure)
- Herniated disc in the past

- Clinically relevant bone marrow edema
- Osteoporosis

Neuropsychiatric

- Serious neurological disorders
  - Stroke
  - Paralysis
  - Multiple sclerosis
  - Convulsions
- Posture insecurity
  - Neurological motor disorders
  - Sensomotoric disorders
  - Vertigo
- Skin sensitivity disorder
- Psychoses
- Dementia
- Drug or alcohol abuse

Oncological

- Metastases of malignant tumors
- Acute malignant disease

Recruitment was initiated through newspaper advertisements in 2 regions of the state of Baden-Wuerttemberg, Germany. Patients interested in participation underwent prescreening by telephone before undergoing an interview and a physical examination. Physical examinations and baseline assessments were conducted by an investigator (HW and KW) and study nurses at an outpatient study center affiliated with the University Hospital Tübingen. All follow-up assessments were conducted remotely via phone calls and questionnaires. Study nurses coordinated follow-up appointments and monitored the completion of follow-up assessments. The trial ended after 12 weeks. No reason for the early termination of the trial was reported throughout the duration of the study. Participants were not paid for trial participation; however, costs resulting directly from trial participation were reimbursed (eg, travel expenses incurred for the baseline visit).

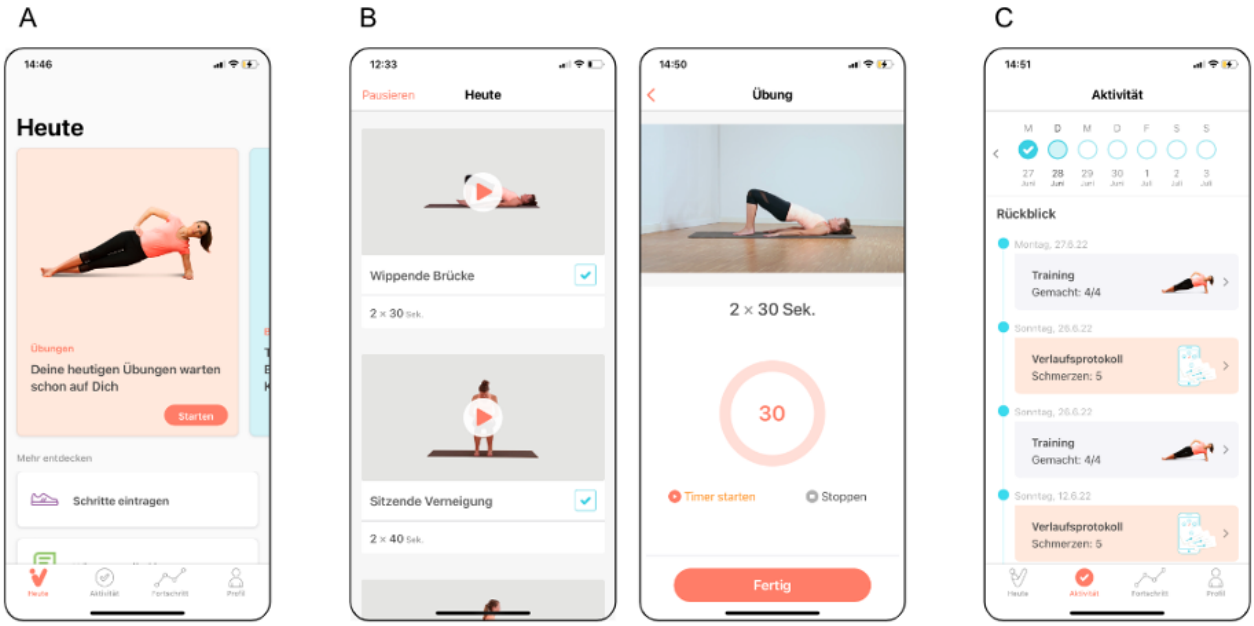
Intervention

The interventional group was provided access to the digital therapeutic on their mobile device free of charge. Patients in this group were asked to exercise at least three days per week throughout the trial period of 12 weeks, and patients were advised to use the default notification setting with a daily reminder displayed as a push notification. The digital therapeutic assessed in this trial was the ViViRA app (ViViRA Health Lab GmbH), an approved DiGA addressing the indication spectrum outlined earlier. It is a medical device used in mobile devices with iOS and Android operating systems, providing a self-directed home exercise program using the principles of movement therapy and functional regional interdependence, as outlined elsewhere [7,14-16]. The intervention is only available through a prescription or an individual subscription. No updates or changes were made to the therapeutic elements of the app during the duration of the study. All patients in the investigational group were provided the same version of the app. The user interface and the prompt of an example exercise is displayed in Figure 1. Patients were prompted to complete 4 exercises per day for 12 consecutive weeks. Guidance on how to exercise is given multimodally using demonstration videos as well as written and audio instructions. After each exercise, patients provide feedback that allows for the continuous adoption of exercise selection based on pain and physical ability. A progression algorithm modifies exercise composition, exercise intensity, and exercise complexity. The development of the progression algorithm was led by an interdisciplinary expert panel consisting of orthopedic surgeons and physiotherapists. The control group received treatment in line with German treatment guidelines [14,17] recommending physiotherapy. This includes physical exercises lasting 15 to 25 minutes guided by a certified physiotherapist. According to the German treatment guidelines [17], treatment includes 6 to 12 such physiotherapy sessions for each prescription. Patients in the control group were assigned to receive physical therapy from a certified physiotherapist of their choice. No influence was exerted on

therapist choice, scheduling, waiting times, or additional physiotherapy sessions; however, any costs incurred were covered by the sponsor of the trial.

Adherence to self-directed exercise therapy was assumed if at least one training was finished per week, and the patient confirmed within the app that they had done the exercises by providing feedback on the feasibility of each exercise and any pain experienced during the exercises. Information on adherence to the standard of care was obtained during follow-up interviews.

Figure 1. Patient interface of the digital home exercise program. (A) Home screen with daily prompt to start the exercises in the German language setting. (B) Composition of 4 exercises based on baseline assessment and patient feedback on pain and functional limitations (left); example of the video- and audio-guided exercise screen (right). (C) Summary of completed exercises, follow-up assessments, and therapeutic progress achieved.



Outcome Measure

Self-reported pain intensity was assessed as the primary outcome measure using a VNRS that was linguistically adapted for German-speaking study participants. A nonequidistant scaling of pain score categories across an 11-point rating scale from 0 to 10 (corresponding to 0-100 mm on a visual analog scale) was used [18]. An adaptation to the proposed scale was made according to Weber et al [19] as the 2 categories for highest pain were integrated. The primary outcome was assessed at baseline and after 2, 6, and 12 weeks in both the control and intervention groups. Secondary analysis of total pain scores and their changes during the study were determined a posteriori. No changes were made to the outcome of the study after it had commenced.

Assessment of Potential Harms

An active surveillance of adverse events (AEs) and unintended effects in both the intervention and control groups was conducted during structured interviews at weeks 2, 6, and 12 after the baseline examination. A differentiation of AE and adverse reactions (ARs) to either the administration of the intervention or control exercise therapy was conducted accordingly.

Sample Size

To determine the required sample size, a trial with a 2-sided question (significance level Cronbach α=5%; power 1-β=80%) was planned. We used retrospective pilot data from patients who had been applying the digital home exercise program between 2018 and 2019. According to the pilot data, the VNRS limit was 1 score point and the approximate SD was 2.5 score points. This produced a standardized delta of Δ=1/2.5=0.4. Calculations with a significance level of Cronbach α=5% and a power of 1-β=80% resulted in 2×99 patients (N=198). To account for a potential dropout to study surveys of approximately 10%, we included 213 patients in this study, with 108 (50.7%) and 105 (49.3%) randomized to the investigation and control groups, respectively. This study was designed to assess the superiority of the intervention against the standard of care treatment.

Randomization

Participants who met the inclusion criteria outlined earlier were randomly assigned to either the investigation or control group. Randomization was based on block randomization with a block size of 6, generated with the SAS module *Proc Plan* The allocation ratio at the time of randomization was 1:1. The randomization list was generated by the data manager of the Contract Research Organization CRM Biometrics GmbH (DS). On entry into the study, each patient was assigned a patient identification number. Using the patient identification number, each included patient was assigned to either the intervention or control group in the sequence specified by the randomization blocks. No deviation from the randomization sequence was reported.

Statistical Analysis

Data analysis was performed by a biostatistician who was not involved in the collection of the analyzed data. Analyses were performed according to the intention-to-treat (ITT) approach. Data from participants with pain scores based on the VNRS at baseline were used in this analysis. It included all participants who were randomized and showed values for the primary variable at baseline. Furthermore, we conducted the same analyses in the prespecified per-protocol (PP) set. This included patients who were not lost to follow-up. In addition, patients who stated at the follow-up assessments that they had received concomitant physiotherapy or taken pain medication during the intervention period were excluded from the PP analysis of the intervention group. Similarly, in the control group, patients who reported concomitant use of pain medications or of a home exercise program during the study period were excluded. Metric data are expressed as means with SDs or 95% CIs. Nominal (sex) and ordinal (shift of pain score) data are reported as the cell frequencies and percentages of patients in each category. Between-group and intragroup differences were calculated using Welch 2-tailed *t* test. Score differences and Cohen delta were calculated for confirmatory treatment group comparisons (intervention vs control group), as well as for intragroup score changes from baseline to follow-ups. Cohen *d* was used for a quantitative and metric-free estimation of the effect size, with values >0.20 defined as small effect sizes, >0.50 as medium effect sizes, and >0.80 as large effect sizes [20]. All hypothesis tests used were 2-sided, and $P \leq .05$ was considered significant. Statistical analyses were performed with SAS, version 94M7 (SAS), and GraphPad Prism, version 9.1.0 (GraphPad).

Ethics Approval

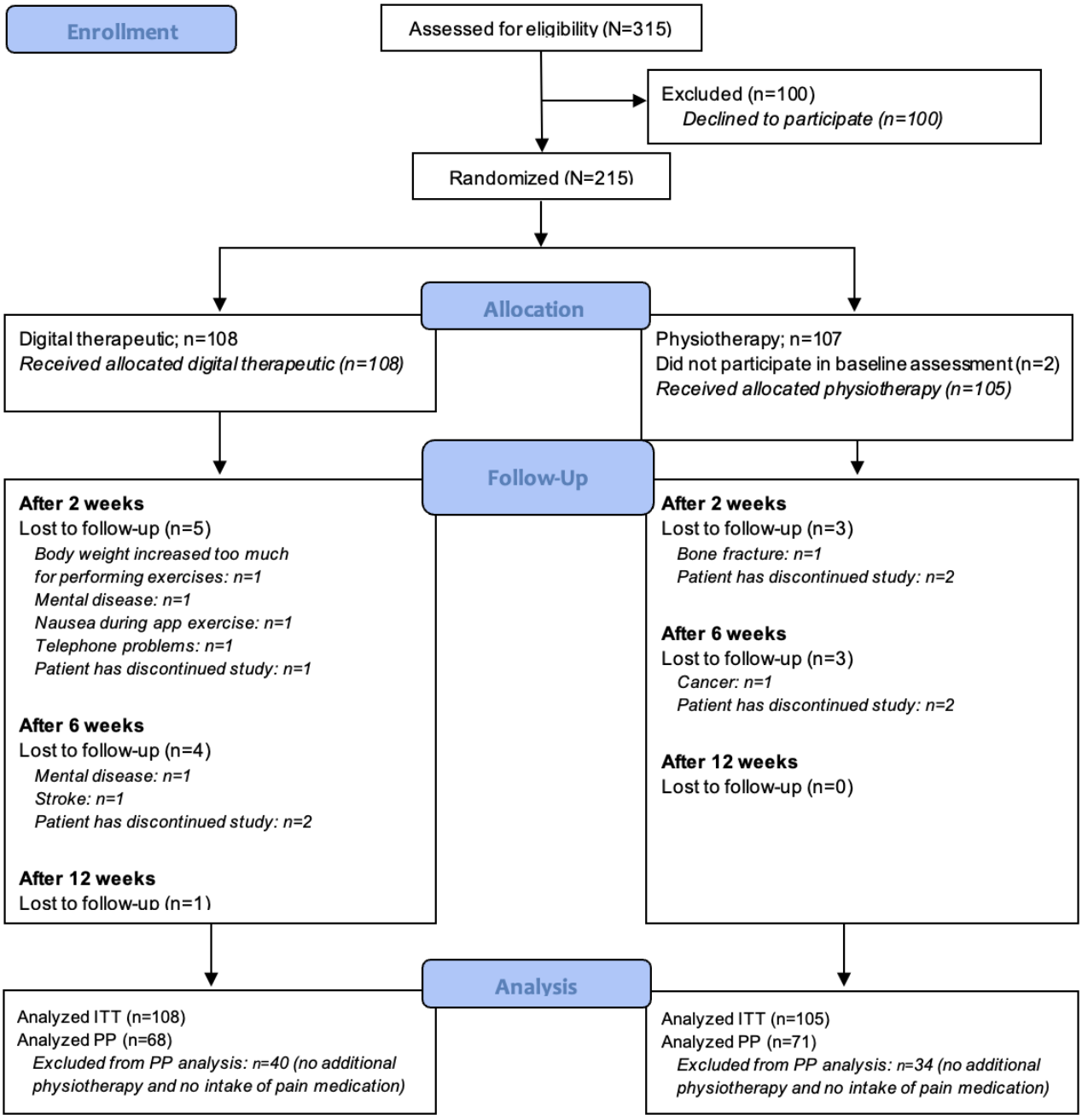
The study concept was reviewed and approved by the Institutional Ethics Commission of the Chamber of Physicians and Surgeons of the State of Baden-Wuerttemberg, Germany, under the registration number F-2020-122 and in agreement with current data protection regulations. The trial is registered at Deutsches Register Klinische Studien (Germany Clinical Trials Register; World Health Organization Primary Register) with the identifier DRKS00022781. Before enrollment in the study, patients received oral information from a trial physician and written patient information that included a description and purpose of the study, possible AEs, the name and address of the insurer, and information on data protection. Thereafter, the patients signed a written informed consent form to participate in the study and consented to the use of their data. This manuscript was prepared in accordance with the 2010 CONSORT (Consolidated Standards of Reporting Trials) guidelines. The intervention studied is outlined in detail in the attached TiDier (template for intervention description and replication) checklist (Multimedia Appendix 1).

Results

Included Patients

A total of 215 patients were enrolled and randomly allocated to the intervention (n=108) and control (n=107) groups. In total, 2 patients in the control group did not respond to the baseline follow-up call and did not provide any outcome data after randomization. Therefore, these patients were considered screening failures and were not included in the subsequent analysis. This reduced the total number of patients in the control group to 105. No violations of the protocol were reported, which would have led to an exclusion from the study. All patients enrolled and randomly allocated were included in the ITT analysis. For the PP analysis, 68 patients of the intervention group and 71 patients of the control group were considered. Figure 2 displays the follow-up chart.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart on screening, inclusion, randomization, follow-up, and analysis. ITT: intention-to-treat; PP: per-protocol.



Recruitment

The recruitment period, as outlined in the Methods section, started in August 2020. The final follow-up was completed by the end of April 2021. The baseline assessment was conducted onsite, whereas follow-up assessments after 2, 6, and 12 weeks were conducted remotely via phone calls and questionnaires. The trial ended after 12 weeks. No reason for an early termination of the trial was reported during the 12-week study period. Table 1 presents the demographic and clinical characteristics of the study participants at the baseline.

Table 1. Baseline demographic and clinical characteristics of the intention-to-treat population (N=213).

	Intervention group[a]	Control group[b]
Participants, n (%)	108 (50.7)	105 (49.3)
Age (years), mean (SD)	57.4 (13.8)	57.3 (13.5)
Sex (female), n (%)	51 (47.2)	62 (59.1)
Indications (ICD-10[c]), n (%)		
M54.4—lumbago with sciatica	44 (40.7)	33 (31.4)
M54.5—low back pain	44 (40.7)	45 (42.9)
M54.9—dorsalgia, unspecified	20 (18.5)	27 (25.7)
Pain score (VNRS[d] 0-10), mean (SD)	6.41 (1.65)	6.05 (1.64)

[a]Patients in the intervention group used a digital home exercise program to treat their back pain.

[b]Patients in the control group received the standard of care (ie, physiotherapy).

[c]ICD-10: International Statistical Classification of Diseases, 10th edition.

[d]VNRS: verbal numerical rating scale.

Primary Outcome

Intragroup Comparison

Pain responses were assessed with a German VNRS validated for nonmalignant pain [18]. All patients were analyzed in the group they were initially assigned to (ie, ITT analysis). From a mean baseline pain score of 6.42 (SD 1.65), the intervention group with 108 participants showed a significant reduction of the pain score to 3.94 (SD 1.79) after 2 weeks to 3.50 (SD 2.21) after 6 weeks and to 3.06 (SD 2.18) after 12 weeks of exercise therapy. These changes are of statistical significance as compared with the assessed baseline pain score (all $P<.001$; Cohen $d>0.8$). Comparing the mean pain scores at baseline and after 12 weeks, the perceived pain decreased by a mean of −3.35 (SD 2.05) score points and −53.1% (SD 29.5%; Table 2). These significantly lower values of reported pain scores at week 12 as compared with the baseline assessment were also found in the PP analysis ($P<.001$).

The control group of 105 participants reported a reduction in pain to a lesser extent as compared with the intervention group. From a reported mean baseline of 6.05 (SD 1.64), a marginal reduction to 5.71 (SD 1.48; $P=.123$; Cohen $d=0.22$) could be observed after 2 weeks. After 6 and 12 weeks, significant pain score reductions to 5.47 (SD 1.80; $P<.05$; Cohen $d=0.34$) and 5.13 (SD 1.91; $P<.001$; Cohen $d=0.52$), respectively, were observed (Figure 3). This reduction in pain corresponds to a mean reduction in perceived pain by −0.91 (SD 1.50) score points and by −14.6% (SD 25.3%; Table 2). Regarding the PP analysis, the described pain reduced significantly ($P<.001$).

Table 2. Absolute and relative pain score (VNRS[a]) changes after 2, 6, and 12 weeks of the intention-to-treat population.

	Intervention group (n=108)			Control group (n=105)		
	2 weeks after baseline	6 weeks after baseline	12 weeks after baseline	2 weeks after baseline	6 weeks after baseline	12 weeks after baseline
Absolute pain score (VNRS) change, mean (SD)	−2.47 (1.74)	−2.92 (2.07)	−3.35 (2.05)	−0.33 (1.42)	−0.58 (1.65)	−0.91 (1.50)
Cohen *d*, between-group comparison[b]	1.35	1.26	1.37	N/A[c]	N/A	N/A
P value[d], between-group comparison	<.001	<.001	<.001	N/A	N/A	N/A
Relative pain score change (%), mean (SD)	−38.0 (22.9)	−45.7 (30.6)	−53.1 (29.5)	−2.45 (24.2)	−7.14 (28.3)	−14.6 (25.3)
Cohen *d*, between-group comparison	1.51	1.31	1.40	N/A	N/A	N/A
P value, between-group comparison	<.001	<.001	<.001	N/A	N/A	N/A

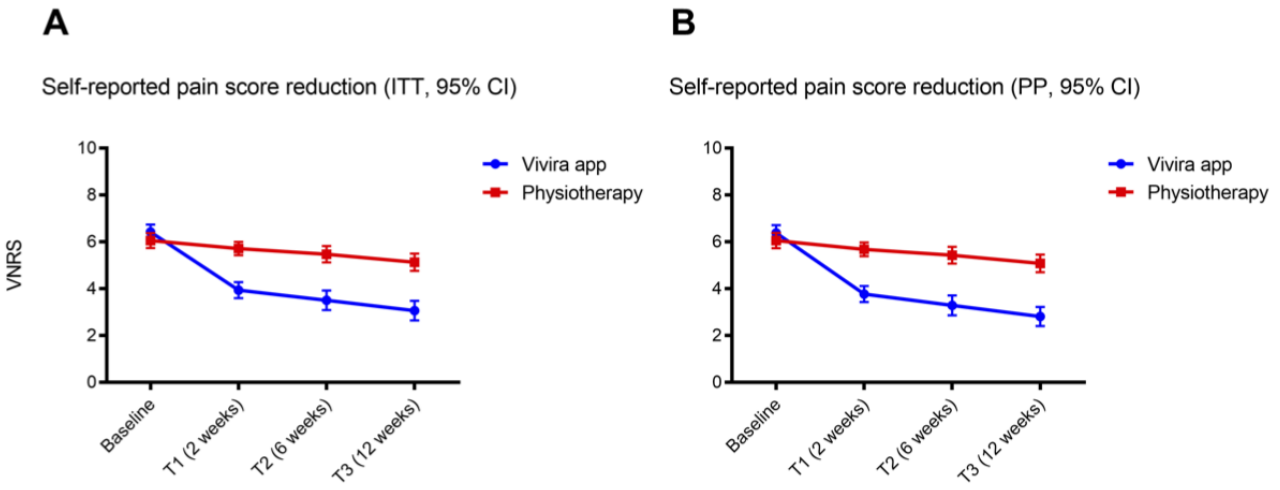
[a]VNRS: verbal numerical rating scale.

[b]The statistical comparison of the between-group differences was calculated using the 2-tailed *t* test.

[c]N/A: not applicable.

[d]A *P* value of <.05 was considered statistically significant.

Figure 3. Pain score values assessed by the verbal numerical rating scale (VNRS) in the (A) intention-to-treat (ITT) and (B) per-protocol (PP) populations at baseline and after 2, 6, and 12 weeks of intervention. Dot plots showing mean pain score values assessed by the VNRS in the patients receiving the digital therapeutic (investigation group, blue) and the conventional physiotherapy (control group, red) of the (A) ITT and (B) PP populations. Error bars indicate 95% CIs.



Between-Group Comparison

Participants of the intervention group reported significantly lower pain intensity than those of the control group from 2 weeks after the start of the study (Figure 2). Between-group differences in reported pain scores showed significantly greater improvements in the intervention group at week 2 (−2.12, 95% CI −2.57 to −1.71; *P*<.01), week 6 (−2.34, 95% CI −2.84 to −1.83; *P*<.01), and week 12 (−2.44, 95% CI −2.92 to −1.95; *P*<.01; Table 3) after the baseline assessment. These results are consistent with those of the PP analysis, where the mean reported pain score was significantly lower in the intervention group as compared with the control group (each *P*<.001; Multimedia Appendix 2).

Table 3. Between-group differences[a] of absolute pain score (verbal numerical rating scale) after 2, 6, and 12 weeks of the intention-to-treat population.

Absolute pain score changes (score points)	Week 2	Week 6	Week 12
Mean	−2.14	−2.34	−2.44
95% CI	−2.57 to −1.71	−2.84 to −1.83	−2.92 to −1.95
P value[b]	<.001	<.001	<.001

[a]The statistical comparison of the between-group differences was calculated using the 2-tailed *t* test.

[b]A *P* value of <.05 was considered statistically significant.

Secondary Analysis of the Overall Pain Scores

A secondary analysis of the pain scores revealed substantially fewer pain exacerbations among the participants of the intervention group as compared with the control group, that is, an increase in pain intensity compared with the reported pain score at baseline.

In the intervention group, 2.77% (3/108) of the patients reported an increase in perceived pain after 6 weeks of treatment but fully recovered after the full duration of the 12-week exercise training program. However, most patients of this group (99/108, 91.7%) reported a reduction in perceived pain.

By contrast, in the control group, 22.9% (24/105) of the patients experienced an increase in pain intensity after 2 weeks and 24.8% (26/105) of the patients after 6 weeks. The number of patients of the control group who reported an increase in pain decreased marginally to 17.1% (18/105) after 12 weeks. However, 60% of the patients of this group reported an improvement in perceived pain (Table 4).

Table 4. Shift of the pain score from baseline to weeks 2, 6, and 12 within the intention-to-treat population.

	Intervention group, VNRS[a] pain score shift			Control group, VNRS pain score shift		
	Improved	No change	Exacerbation	Improved	No change	Exacerbation
Week 2, n (%)	97 (89.8)	11 (10.2)	0 (0)	42 (40)	39 (37.1)	24 (22.9)
Week 6, n (%)	95 (87.9)	19 (9.3)	3 (2.8)	50 (47.6)	29 (27.6)	26 (24.8)
Week 12, n (%)	99 (91.7)	9 (8.3)	0 (0)	63 (60.0)	24 (22.9)	18 (17.1)

[a]VNRS: verbal numerical rating scale.

Adherence

Patients in the intervention group of the ITT cohort were using the exercise therapy on an average of 5.77 days out of 7 possible days per week. This corresponds to an adherence rate of 89.9%. The patients of the control group received a mean of 6.94 (SD 2.94) physiotherapy sessions during the 12-week study period. Adherence in the control group was defined as the percentage of 695 physiotherapy sessions completed versus the planned number of 714 sessions, resulting in an adherence rate of 97.3%.

Adverse Reactions and Adverse Events

ARs were reported by 34.3% (37/108) of the patients in the intervention group and 29.5% (31/105) of the patients in the control group and are outlined in Multimedia Appendix 3. None of the reported ARs led to the discontinuation of the intervention in either group. In addition, no serious AEs were reported. No privacy breaches or substantial technical problems were detected.

Discussion

Principal Findings

This study aimed to assess the comparative effectiveness of a digital home exercise program and the standard of care treatment for unspecific and degenerative back pain. The results of the present study show that the use of a digital home exercise program can lead to a significant and clinically relevant reduction in patient-reported unspecific and degenerative back pain. Moreover, the results of the present study indicate that the reduction of self-reported pain intensity achievable with the digital therapeutic under investigation is superior to the reduction of self-reported pain intensity achieved with the standard of care (mean difference of the assessed pain score at 12 weeks: −2.44, 95% CI −2.92 to −1.95, in favor of the intervention group).

Limitations

The data presented in this paper contribute to the growing body of knowledge in the field of digital therapeutic interventions. Through a pragmatic randomized controlled design, this trial aimed to substantiate the evidence for the effectiveness of the digital home exercise program ViViRA. Nonetheless, we see factors that limit the external validity of our study and thus the generalizability of our findings.

First, the decentralized nature of digital therapeutics is a key factor leading to better access to and availability of therapeutic resources compared with physiotherapy treatments in a physiotherapist's practice. This comes at the expense of a close interpersonal relationship between patients and health care professionals, which naturally contributes to the effectiveness of therapeutic interventions [21,22]. However, as this study relied on the conventional (ie, out of app) collection of data through phone calls and questionnaires (as compared with in-app and real-world data analyses), the trial staff maintained close contact with the enrolled patients. Therefore, a potential observer bias as well as a detection bias needs to be taken into account when applying these results to a real-world use scenario, and further research on observational or real-world use data is required to assess the extent of these potential biases.

Second, the enrollment for the trial presented was primarily based on newspaper advertisements in 2 regions in the German federal state of Baden-Wuerttemberg. As this differs from the enrollment in clinical practice, a selection bias is plausibly present in this study, as a DiGA typically requires a prescription from a health care professional (ie, provider-driven initiation of therapy) and is only very limited accessible through self-selection (ie, patient-driven initiation of therapy). As discussed earlier, more research on the relevance of these differential patient motivations is required.

Third, the availability of physiotherapists is limited and varies by region. Therefore, system-related waiting times are likely to reduce the therapeutic density (ie, the number of therapeutic sessions per week), which affects the expected effectiveness in the control group. This difference in effectiveness between the intervention and the standard of care is likely to be emphasized by the decentralized and on-demand availability of digital therapeutics. This could explain the small effect of physiotherapy on pain intensity in comparison with the intervention examined. However, given that improved access to and availability of digital therapeutics are key characteristics of digital therapeutics, we considered the comparison appropriate in the context of real-world use.

Further limitations of this study include the nonblinded design, which was required as the mode of administration of the intervention and the control differed significantly and could not be feasibly blinded, and the lack of an objective measure for perceived pain intensity, which is challenging because of its highly individual nature. Generally, self-reported pain intensity is considered to be validly measured by different pain scales. This study relied on a German VNRS that has been validated for nonmalignant pain [18]. In addition, sufficient comparability with other unidimensional pain intensity scales has been demonstrated by other researchers [23,24].

Comparison With Prior Work

This study focused on assessing the effectiveness of digital therapy compared with physiotherapy, the standard of care. During the 12-week exercise program with digital therapeutic, 91.7% (99/108) of the patients described pain relief. On the basis of the VNRS, which has been used to quantify pain, this corresponds to a mean pain relief of −52.3%. These results complement the existing literature, as comparable positive effects on pain intensity between −33.3% and −81% have been described in several studies assessing the effectiveness of digital therapeutics for musculoskeletal conditions [25-29]. As equivalent intervention periods of 12 weeks were studied, we deemed the comparison with these studies applicable. However, we deem the comparison with the work of Shebib et al [27] as particularly comparable, as this group also assessed a stand-alone intervention for the treatment of lower back pain. This group demonstrated an average pain score improvement between 52% and 64% [27]. Other works, for example, from Priebe et al [26] and Sandal et al [29], pursued an add-on approach to augment the existing infrastructure of care and can, therefore, not be considered a stand-alone intervention. Apart

from the effectiveness of the digital therapeutic, it is also necessary to consider the effects that were achieved in the control group by the standard of care. Interestingly, similar to the previously mentioned studies [25-27], the extent to which physiotherapy was able to achieve a substantial reduction in pain intensity was lower as compared with digital exercise therapy. As outlined earlier, we do not interpret this as evidence for physiotherapy not being effective in reducing self-reported pain intensity but attribute this primarily to the different modes of administration. The centralized (ie, onsite) and synchronous (ie, by appointment) administration of physiotherapy limits the patient-specific adaptation of therapy intensity and frequency and, hence, can lead to suboptimal therapeutic results. We see this reflected in the average therapy frequency in our data: patients in the intervention group used the exercise training on average for 5.77 days per week. In comparison, patients in the control group received 6.94 physiotherapy sessions during the entire 12-week study period. In addition, the German health care system has reimbursement limits and provider-specific budgets for the number of physical therapy sessions available to a patient. These system-inherent limitations, from our point of view, reduce the achievable positive effects of conventional physiotherapy, as observed in this study. By contrast, no such differences or much smaller differences between a digital therapeutic and the standard of care treatment have been observed in other studies. Sandal et al [29], for example, described significant, though smaller, between-group differences in pain intensity in favor of an artificial intelligence-based app to self-management support system for treatment of lower back pain, compared with standard of care. However, as discussed earlier, the overall approach of this group differs from the assessment presented here, as Sandal et al [29] studied an add-on intervention for the treatment of lower back pain. Similarly, Koppenaal et al [30] assessed the effectiveness of blended physiotherapy (digital exercise training with face-to-face physiotherapy sessions) compared with the standard of care and found no group differences in pain reduction. However, an exception to this overall finding is the group of patients at a high risk of developing persistent low back pain in which blended therapy was superior to physiotherapy in terms of average reported pain reduction [30]. These results underscore, from our perspective, the advantages of decentralized and immediately available digital therapies for the treatment of back pain. Furthermore, Lara-Palomo et al [31] found no difference in effectiveness in reducing back pain when comparing digital health apps and standard face-to-face care in a systematic review and meta-analysis. Although the available evidence was only considered to be of moderate quality, we interpret these results as yet another indicator for the quality and effectiveness of care that can be delivered through digital therapeutics.

Finally, and especially in the case of patient-oriented outcomes, it is important to assess the clinical relevance of the results obtained. Several thresholds for assessing the clinical relevance of improvements in pain intensity have been defined in the literature. Exemplarily, and according to Ostelo et al [32], this threshold is a reduction of 30%, considered a minimally important change, while Holdgate et al [33] referred to a 1.4 score point improvement in VNRS as the minimum clinically significant difference. Applying these criteria underlines that the effect of the digital home exercise program on pain intensity was not only statistically significant but also clinically relevant for patients. The positive effect on a reduction in pain intensity to a clinically relevant extent was measurable after the second week of exercise therapy in the investigation group. Interestingly, the thresholds discussed above were not met by patients in the control group. To our knowledge, none of the previous studies with a digital therapeutic focused and described such an early effect. Before this background, digital home exercise programs can be considered a veritable therapeutic option for unspecific and degenerative back pain, which is in line with national and international treatment recommendations [6,14,15] that prioritize movement and exercise therapy over medication and more invasive therapeutic measures. In terms of potential harms associated with the use of the digital therapeutic, we noted several AEs, all of which were transient in nature (Multimedia Appendix 2). Therefore, we conclude that no intolerable risks are associated with the use of the program assessed within the scope of its approved indications and considering the exclusion criteria.

Conclusions

In the face of an increasing burden of disease from unspecific and degenerative musculoskeletal conditions, novel and innovative therapeutic approaches are required to ensure access to and availability of effective care for this spectrum of conditions. With the introduction of the DiGA into the collectively funded German health care system, a regulatory framework for the system-wide implementation of digital therapeutics was created. This study presents effectiveness data for one of the first fully approved DiGAs and shows significant and clinically relevant improvements in self-reported pain intensity. These improvements were superior to those of the control group, representing the current standard of care in the German health care system. By expanding the available therapeutic capacities for unspecific and degenerative back pain through a decentralized and on-demand digital therapeutic, a significant added value in pain management can be achieved.

Given the high burden of disease caused by back pain and the limited availability of and access to adequate health care, digital apps are an efficient treatment option for unspecific and degenerative back pain. In view of this, replication of the present trial in further independent studies considering additional outcome parameters, such as function, with longer follow-up periods, for example, 6 and 12 months, and its applicability in other countries and health care systems is of great interest. In addition, particularly in the field of digital therapeutics, further research on available real-world use data will complement the formalized and trial-based assessments of such therapeutics. By generating an increasing body of evidence as well as integrating digital apps into health care systems, digital therapeutics can contribute significantly to health care in the indication area.

Acknowledgments

The study was sponsored by ViViRA Health Lab GmbH. The sponsor provided access to the digital home exercise program during the trial. The authors acknowledge the work of Markus Klingenberg, who developed the therapy concept of the medical software device assessed in this research. This includes the digital implementation of the functional therapeutic approach, the device's software-patient feedback interface, and its exercise progression algorithm.

Authors' Contributions

KW and HW contributed to the study conception and design; BZ and HW contributed to the study performance; BZ contributed to the generation of the data; KW, HW, and LB contributed to data analysis; and MB and DS performed the statistical analysis. All authors contributed to the writing and editing of the paper.

Conflicts of Interest

HW, BZ, MB, DS, and KW were responsible for devising the study design and overseeing the study and data analysis. They are researchers, clinicians, and statisticians who are independent of ViViRA Health Lab GmbH. They received salaries (BZ, MB, and DS) or honoraria (HW and KW) for their involvement in the study. BS and LB are employed by ViViRA Health Lab GmbH.

Multimedia Appendix 1
Template for intervention description and replication (TIDieR) checklist for the intervention studied.
[DOCX File , 77 KB - jmir_v24i10e41899_app1.docx]

Multimedia Appendix 2
Pain score values assessed by the verbal numerical rating scale in the per-protocol population at baseline and after 2, 6, and 12 weeks of intervention.
[DOCX File , 13 KB - jmir_v24i10e41899_app2.docx]

Multimedia Appendix 3
Adverse reactions reported by the study population during the 12-week intervention period.
[DOCX File , 13 KB - jmir_v24i10e41899_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2070 KB - jmir_v24i10e41899_app4.pdf]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Woolf AD, Erwin J, March L. The need to address the burden of musculoskeletal conditions. Best Pract Res Clin Rheumatol 2012 Apr;26(2):183-224. [doi: 10.1016/j.berh.2012.03.005] [Medline: 22794094]
3. Speerin R, Slater H, Li L, Moore K, Chan M, Dreinhöfer K, et al. Moving from evidence to practice: models of care for the prevention and management of musculoskeletal conditions. Best Pract Res Clin Rheumatol 2014 Jun;28(3):479-515. [doi: 10.1016/j.berh.2014.07.001] [Medline: 25481427]
4. Araujo de Carvalho I, Epping-Jordan J, Pot AM, Kelley E, Toro N, Thiyagarajan JA, et al. Organizing integrated health-care services to meet older people's needs. Bull World Health Organ 2017 Nov 01;95(11):756-763 [FREE Full text] [doi: 10.2471/BLT.16.187617] [Medline: 29147056]
5. Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010 Dec;19(12):2075-2094 [FREE Full text] [doi: 10.1007/s00586-010-1502-y] [Medline: 20602122]
6. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Jun 09;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
7. Chiarotto A, Koes BW. Nonspecific low back pain. N Engl J Med 2022 May 05;386(18):1732-1740. [doi: 10.1056/NEJMcp2032396] [Medline: 35507483]
8. Briggs AM, Shiffman J, Shawar YR, Åkesson K, Ali N, Woolf AD. Global health policy in the 21st century: challenges and opportunities to arrest the global disability burden from musculoskeletal health conditions. Best Pract Res Clin Rheumatol 2020 Oct;34(5):101549 [FREE Full text] [doi: 10.1016/j.berh.2020.101549] [Medline: 32713802]

9. Crettenden I, Poz MD, Buchan J. Right time, right place: improving access to health service through effective retention and distribution of health workers. Hum Resour Health 2013 Nov 25;11:60 [FREE Full text] [doi: 10.1186/1478-4491-11-60] [Medline: 24274820]
10. Buchan J. Reviewing the benefits of health workforce stability. Hum Resour Health 2010 Dec 14;8:29 [FREE Full text] [doi: 10.1186/1478-4491-8-29] [Medline: 21156050]
11. Essery R, Geraghty AW, Kirby S, Yardley L. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil 2017 Mar;39(6):519-534. [doi: 10.3109/09638288.2016.1153160] [Medline: 27097761]
12. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
13. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
14. Bundesärztekammer, Kassenärztlicher Bundesvereinigung, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Nationale VersorgungsLeitlinie (NVL) Nicht-spezifischer Kreuzschmerz, 2. Auflage. Nicht-spezifischer Kreuzschmerz. 2017. URL: https://www.leitlinien.de/themen/kreuzschmerz [accessed 2022-08-12]
15. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
16. Mueller J, Niederer D. Dose-response-relationship of stabilisation exercises in patients with chronic non-specific low back pain: a systematic review with meta-regression. Sci Rep 2020 Oct 09;10(1):16921 [FREE Full text] [doi: 10.1038/s41598-020-73954-9] [Medline: 33037280]
17. Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Heilmitteln in der vertragsärztlichen Versorgung. Gemeinsamer Bundesausschuss. 2021. URL: https://www.g-ba.de/downloads/62-492-2717/HeilM-RL_2021-10-21_iK-2022-01-22.pdf [accessed 2022-08-12]
18. Aicher B, Peil H, Peil B, Diener HC. Pain measurement: Visual Analogue Scale (VAS) and Verbal Rating Scale (VRS) in clinical trials with OTC analgesics in headache. Cephalalgia 2012 Feb;32(3):185-197. [doi: 10.1177/03331024111430856] [Medline: 22332207]
19. Weber M, Schüz J, Kuball J, Gamm H, Jage J. [Pain assessment in invasive diagnostic procedures. Comparison of an eleven-point numerical rating scale and a six-point verbal rating scale for pain measurement in bone marrow puncture]. Schmerz 2005 Nov;19(6):513-519. [doi: 10.1007/s00482-005-0393-z] [Medline: 15815930]
20. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
21. Hush JM, Cameron K, Mackey M. Patient satisfaction with musculoskeletal physical therapy care: a systematic review. Phys Ther 2011 Jan;91(1):25-36. [doi: 10.2522/ptj.20100061] [Medline: 21071504]
22. Kinney M, Seider J, Beaty AF, Coughlin K, Dyal M, Clewley D. The impact of therapeutic alliance in physical therapy for chronic musculoskeletal pain: a systematic review of the literature. Physiother Theory Pract 2020 Aug;36(8):886-898. [doi: 10.1080/09593985.2018.1516015] [Medline: 30265840]
23. Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. Acad Emerg Med 2001 Dec;8(12):1153-1157 [FREE Full text] [doi: 10.1111/j.1553-2712.2001.tb01132.x] [Medline: 11733293]
24. Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, European Palliative Care Research Collaborative (EPCRC). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011 Jun;41(6):1073-1093 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.08.016] [Medline: 21621130]
25. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May 3;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
26. Priebe JA, Haas KK, Moreno Sanchez LF, Schoefmann K, Utpadel-Fischler DA, Stockert P, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020;13:1823-1838 [FREE Full text] [doi: 10.2147/JPR.S260761] [Medline: 32765057]
27. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019 Jan 7;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
28. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
29. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021 Oct 01;181(10):1288-1296 [FREE Full text] [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]

30. Koppenaal T, Pisters MF, Kloek CJ, Arensman RM, Ostelo RW, Veenhof C. The 3-month effectiveness of a stratified blended physiotherapy intervention in patients with nonspecific low back pain: cluster randomized controlled trial. J Med Internet Res 2022 Feb 25;24(2):e31675 [FREE Full text] [doi: 10.2196/31675] [Medline: 35212635]
31. Lara-Palomo IC, Gil-Martínez E, Ramírez-García JD, Capel-Alcaraz AM, García-López H, Castro-Sánchez AM, et al. Efficacy of e-health interventions in patients with chronic low-back pain: a systematic review with meta-analysis. Telemed J E Health (forthcoming) 2022 May 09. [doi: 10.1089/tmj.2021.0599] [Medline: 35532971]
32. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008 Jan 01;33(1):90-94. [doi: 10.1097/BRS.0b013e31815e3a10] [Medline: 18165753]
33. Holdgate A, Asha S, Craig J, Thompson J. Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain. Emerg Med (Fremantle) 2003;15(5-6):441-446. [doi: 10.1046/j.1442-2026.2003.00499.x] [Medline: 14992058]

Abbreviations

AE: adverse event
AR: adverse reaction
BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte
CONSORT: Consolidated Standards of Reporting Trials
DiGA: Digitale Gesundheitsanwendungen
ITT: intention-to-treat
PP: per-protocol
TiDier: template for intervention description and replication
VNRS: verbal numerical rating scale

Edited by R Kukafka, G Eysenbach; submitted 13.08.22; peer-reviewed by A Chiarotto, M Hong; comments to author 07.09.22; revised version received 25.09.22; accepted 08.10.22; published 28.10.22.

Please cite as:
Weise H, Zenner B, Schmiedchen B, Benning L, Bulitta M, Schmitz D, Weise K
The Effect of an App-Based Home Exercise Program on Self-reported Pain Intensity in Unspecific and Degenerative Back Pain: Pragmatic Open-label Randomized Controlled Trial
J Med Internet Res 2022;24(10):e41899
URL: https://www.jmir.org/2022/10/e41899
doi:10.2196/41899
PMID:36215327

Original Paper

Racial and Ethnic Differences in Outcomes of a 12-Week Digital Rehabilitation Program for Musculoskeletal Pain: Prospective Longitudinal Cohort Study

Justin Scheer[1*], MD; Fabíola Costa[2*], PhD; Maria Molinos[2], PhD; Anabela Areias[2], PhD; Dora Janela[2], PT; Robert G Moulder[3], PhD; Jorge Lains[4,5], MD; Virgílio Bento[2], PhD; Vijay Yanamadala[2,6,7], MBA, MD; Steven P Cohen[8,9,10,11,12,13], MD, PhD; Fernando Dias Correia[2,14], MD, PhD

[1]Department of Neurological Surgery, University of California, San Francisco, CA, United States
[2]Sword Health, Inc, Draper, UT, United States
[3]Institute for Cognitive Science, University of Colorado Boulder, Boulder, CO, United States
[4]Faculty of Medicine, Coimbra University, Coimbra, Portugal
[5]Rovisco Pais Medical and Rehabilitation Centre, Tocha, Portugal
[6]Department of Surgery, Frank H Netter School of Medicine, Quinnipiac University, Hamden, CT, United States
[7]Department of Neurosurgery, Hartford Healthcare Medical Group, Westport, CT, United States
[8]Department of Anesthesiology & Critical Care Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States
[9]Department of Anesthesiology, Uniformed Services University of the Health Sciences, Bethesda, MD, United States
[10]Department of Physical Medicine and Rehabilitation, Johns Hopkins School of Medicine, Baltimore, MD, United States
[11]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, United States
[12]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, United States
[13]Department of Physical Medicine and Rehabilitation, Uniformed Services University of the Health Sciences, Bethesda, MD, United States
[14]Neurology Department, Centro Hospitalar e Universitário do Porto, Porto, Portugal
[*]these authors contributed equally

Corresponding Author:
Fernando Dias Correia, MD, PhD
Sword Health, Inc
13937 Sprague Lane, Suite 100
Draper, UT, 84020
United States
Phone: 1 385 308 8034
Fax: 1 801 206 3433
Email: fcorreia@swordhealth.com

Abstract

Background: Musculoskeletal (MSK) pain disproportionately affects people from different ethnic backgrounds through higher burden and less access to care. Digital care programs (DCPs) can improve access and help reduce inequities. However, the outcomes of such programs based on race and ethnicity have yet to be studied.

Objective: We aimed to assess the impact of race and ethnicity on engagement and outcomes in a multimodal DCP for MSK pain.

Methods: This was an ad hoc analysis of an ongoing decentralized single-arm investigation into engagement and clinical-related outcomes after a multimodal DCP in patients with MSK conditions. Patients were stratified by self-reported racial and ethnic group, and their engagement and outcome changes between baseline and 12 weeks were compared using latent growth curve analysis. Outcomes included program engagement (number of sessions), self-reported pain scores, likelihood of surgery, Generalized Anxiety Disorder 7-item scale, Patient Health Questionnaire 9-item, and Work Productivity and Activity Impairment. A minimum clinically important difference (MCID) of 30% was calculated for pain, and multivariable logistic regression was performed to evaluate race as an independent predictor of meeting the MCID.

Results: A total of 6949 patients completed the program: 65.5% (4554/6949) of them were non-Hispanic White, 10.8% (749/6949) were Black, 9.7% (673/6949) were Asian, 9.2% (636/6949) were Hispanic, and 4.8% (337/6949) were of other racial or ethnic

backgrounds. The population studied was diverse and followed the proportions of the US population. All groups reported high engagement and satisfaction, with Hispanic and Black patients ranking first among satisfaction despite lower engagement. Black patients had a higher likelihood to drop out (odds ratio [OR] 1.19, 95% CI 1.01-1.40, *P*=.04) than non-Hispanic White patients. Hispanic and Black patients reported the highest level of pain, surgical intent, work productivity, and impairment in activities of daily living at baseline. All race groups showed a significant improvement in all outcomes, with Black and Hispanic patients reporting the greatest improvements in clinical outcomes. Hispanic patients also had the highest response rate for pain (75.8%) and a higher OR of meeting the pain MCID (OR 1.74, 95% CI 1.24-2.45, *P*=.001), when compared with non-Hispanic White patients, independent of age, BMI, sex, therapy type, education level, and employment status. No differences in mental health outcomes were found between race and ethnic groups.

Conclusions: This study advocates for the utility of a DCP in improving access to MSK care and promoting health equity. Engagement and satisfaction rates were high in all the groups. Black and Hispanic patients had higher MSK burden at baseline and lower engagement but also reported higher improvements, with Hispanic patients presenting a higher likelihood of pain improvement.

(J Med Internet Res 2022;24(10):e41306) doi:10.2196/41306

KEYWORDS

physical therapy; telerehabilitation; digital therapy; eHealth; telehealth; musculoskeletal conditions; race; ethnicity; pain; diversity; equity; mobile phone

Introduction

Musculoskeletal (MSK) pain affects approximately 1.71 billion people worldwide [1] and up to 83% of those seeking medical care through ambulatory visits [2]. MSK pain results in significant disability and suffering, with a cost of up to US $465 billion in total medical expenditure in 2019 in the United States [3]. Exercise-based physical therapy is the mainstay of treatment for more invasive strategies such as surgery [2,4-6]. However, poor treatment adherence is a barrier to successful treatment [7-9]. Adherence may be affected by a number of factors, such as lack of (1) motivation or self-discipline, (2) provider availability or long waiting list, (3) available time or long distances to travel, and (4) social distancing and concern for contracting an illness around other people [8,10,11].

A new era of telehealth, specifically digital physical therapy, has recently emerged and been brought to the forefront of the COVID-19 pandemic [11]. These digital programs have shown great promise in treating a wide range of MSK pain disorders [9,12,13] and are feasible and effective compared with traditional physical therapy [14-20]. Digital therapy can increase access to care by reducing travel limitations and time barriers and eliminating geographic restrictions. It can also increase adherence by allowing patients to work at their own pace on their own time, thereby increasing empowerment and self-management [7,9].

Despite the many benefits of telehealth, inequities remain based on age, income, health education, digital literacy, and English proficiency [21-23]. Individuals with limited digital literacy or access to technology may not have the means to engage in a digital care program (DCP) [23]. In addition, one major reason for inequities in health care, particularly in telehealth and physical therapy, is race and ethnicity [24-27]. People from racial and ethnic minority groups have been reported to experience higher levels of pain and disability [28,29]. In fact, it is known that pain is not equally experienced among different racial and ethnic groups [24,25,30,31].

Weber et al [32] reported that Black and Hispanic patients were more likely to go to the emergency room or an in-person visit than use telehealth [32]. Other studies have reported similar results, with patients from racial and ethnic minority groups not accessing telehealth as much as non-Hispanic White patients [21,27,33]. Moreover, these populations have been shown to have worse outcomes following rehabilitation than non-Hispanic White patients [25,28].

To our knowledge, no study has been conducted on the impact of race and ethnicity on engagement and outcomes following telerehabilitation for MSK pain. Previously, we have reported clinical studies with a multimodal DCP that combined exercise-based physical therapy with psychoeducational components via a comprehensive approach to pain management [17-19,34,35]. Similar results on pain and functionality were observed with this DCP compared with in-person approaches in patient rehabilitation after surgery, both in the short and long term [17-19,36]. The purpose of this study was to assess the impact of racial and ethnic differences on engagement and outcomes in a completely remote, multimodal DCP for MSK pain with the hypothesis that all races would engage similarly and experience significant improvement in outcomes following the program.

Methods

Study Design

This study was an ad hoc analysis of an ongoing decentralized single-arm clinical trial investigating engagement and clinical-related outcomes after multimodal DCP in patients with MSK conditions. The home-based DCP was delivered between June 29, 2020, and May 26, 2022.

Ethics Approval

The trial was prospectively registered at ClinicalTrials.gov (NCT04092946) on September 17, 2019, and approved by the New England Institutional Review Board (number 120190313) on June 18, 2020.

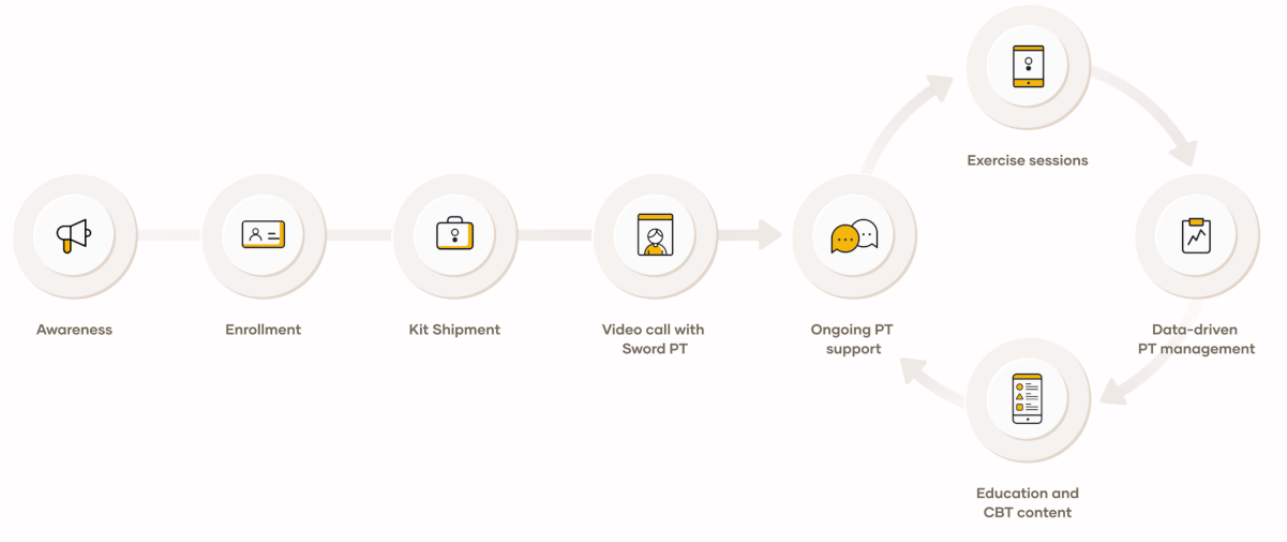
Population

Adults (aged ≥18 years) from 50 states and the District of Columbia in the US beneficiaries of health plans covering the Sword Health program and reporting chronic MSK pain (>12 weeks in the spine, upper, or lower limbs) were eligible to apply to Sword Health's (Draper, Utah, United States) DCP. Employees and their dependents were notified of their eligibility by their employer via email and on-site events and enrolled on the web for free through a dedicated website. During the enrollment phase, all participants were educated about the program and asked to provide informed consent to participate in the clinical trial. All participants completed a baseline form providing demographic data and details regarding their clinical condition, alongside specific questions to screen for potential clinical red flags, which were posteriorly assessed by an assigned physical therapist (PT) through an onboarding video call. The exclusion criteria were as follows: (1) a health condition (eg, cardiac or respiratory) incompatible with at least 20 minutes of light-to-moderate exercise; (2) receiving treatment for active cancer; and (3) reporting any of the following signs and symptomatology, rapidly progressive loss of strength, numbness in either the arms or legs, unexplained changes in bowel or urinary function in the previous 2 weeks.

Intervention

DCP has been previously described elsewhere [17-19,34,35]. The program consisted of a 12-week digitally delivered intervention that included exercise, education, and cognitive behavioral therapy (CBT). The participant journey during the DCP is depicted in Figure 1. Upon registration on the website, a condition-specific kit is shipped corresponding to a Food and Drug Administration–listed class II medical device that comprises inertial motion trackers, a mobile app on a dedicated tablet, and a cloud-based portal.

Figure 1. Participant journey during the digital care program. CBT: cognitive behavioral therapy; PT: physical therapist.



An onboarding call with an assigned PT is scheduled, which is then responsible for program tailoring (according to the specific condition) and monitoring. Personalized exercise sessions were performed independently at the patients' convenience (at least three sessions per week were recommended). In case of a lack of internet access at home, a Wi-Fi hotspot was provided. Exercises were displayed on the tablet, with trackers allowing real-time video and audio biofeedback on performance. A cloud-based portal stored data related to exercise sessions (adherence, existence or absence of movement errors, and level of pain and fatigue during exercises), which enabled asynchronous and remote monitoring and adjustment by the assigned PT. The educational content provided was condition-specific, whereas CBT was general MSK pain-oriented. The educational component of the program was developed according to current clinical guidelines and research and included topics focused on anatomy, physiology, symptoms, evidence-based treatments, fear avoidance, and active coping skills (including dealing with feelings of anxiety and depression). The CBT program was based on mindfulness, acceptance and commitment therapy, empathy-focused therapy, fear-avoidance behavior, and constructive coping. Education and CBT materials were delivered to the patients weekly through written articles, audio content, and interactive modules. Bidirectional communication with the assigned PT was ensured through a built-in secure chat within the smartphone app and video calls. Participants were considered dropouts if they did not engage in any exercise sessions for 28 consecutive days. Participants were included if they were compliant with the intervention but failed to complete a given reassessment survey.

Demographic Data

Demographic data collected included age, race, MSK condition, BMI, sex, educational level, and employment status. The race and ethnic groups included Asian, Black, Hispanic, other, and non-Hispanic White. The gender category included men, women, nonbinary, and "prefer not to specify." A total of 8 educational levels were collected and then grouped as high school or less, some college including bachelor's degree, and some graduate school including master's and doctorate degrees. Furthermore, 8 employment status categories were collected and grouped as employed or not employed.

Clinical Outcomes

Outcomes were collected at baseline and at 4, 8, and 12 weeks, and the mean changes were calculated between baseline and 12 weeks. These included the following:

1. Patient engagement was measured as follows: (1) completion of the program (considered as the retention rate), (2) total number of completed exercise sessions over the 12 weeks, (3) total time spent performing exercise sessions, (4) mean number of sessions per week, (5) total articles read, (6) total interactions with the PT, and (7) overall satisfaction through the question: "On a scale from 0 to 10, how likely is it that you would recommend this intervention to a friend or neighbor?"
2. Pain, using the Numerical Pain Rating Scale, through the question "Please rate your average pain over the last 7 days" from 0 (no pain at all) to 10 (worst pain imaginable)". A minimum clinically important difference (MCID) of 30% between the baseline and treatment end was calculated and analyzed [37,38].
3. Willingness to undergo surgery: "How likely are you to have surgery to address your condition in the next 12 months?" (range: 0—not at all likely; 100—extremely likely).
4. Generalized Anxiety Disorder 7-item scale (GAD-7; range 0-21) [39] was used to assess anxiety, and Patient Health Questionnaire 9-item (PHQ-9; range 0-27) to assess depression [40]. Higher scores indicated worse symptoms.
5. The Work Productivity and Activity Impairment (WPAI) for general health questionnaire evaluated overall work impairment in employed participants (WPAI overall: total presenteeism and absenteeism from work), presenteeism (WPAI work), absenteeism (WPAI time), and activity impairment (WPAI activity) [41]. Higher scores indicated greater impairment.

Safety and Adverse Events

Patients were advised to report any adverse events to the dedicated PT through available communication channels for further assessment.

Statistical Analysis

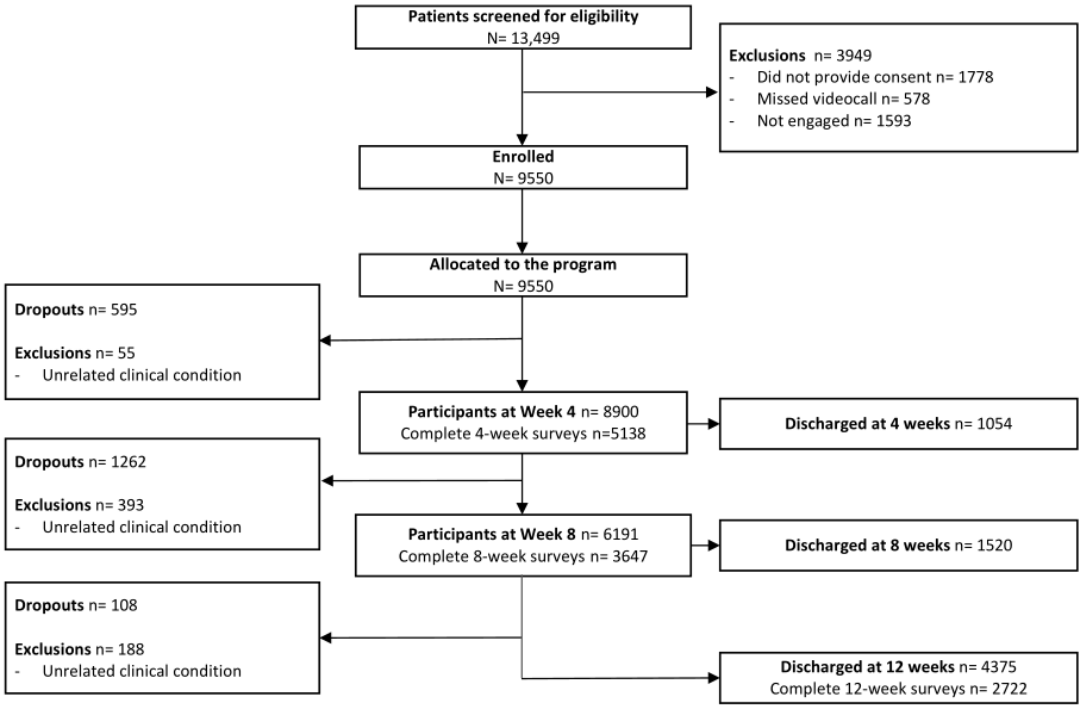
A descriptive analysis of the study population demographics (age, BMI, gender, education level, and employment status), clinical data, and engagement metrics was performed. Patients who completed the 12-week program were defined as "completers" and those that did not were defined as "noncompleters." Statistical analysis between completers and noncompleters was performed using the 2-sample independent *t* test, Mann-Whitney *U* test, 1-way ANOVA with Bonferroni post hoc, or chi-square test.

Latent growth curve analysis (LGCA) was used to estimate trajectories of outcome variables over time, as previously described [34]. The analysis was performed following both an intention-to-treat and a per-protocol approach. Advantages of using LGCA include providing a measure of fitness and addressing missing data through full information maximum likelihood, which outperforms other modern imputation models, such as multiple imputation by chained equations or listwise deletion [42]. The model was adjusted for age, gender, and BMI and fitted as a random effect. Subpopulations were analyzed by filtering cases at baseline: GAD-7, PHQ-9 ≥5 points [39,40], and surgery intention and WPAI (overall, work, time, and activity) >0 points. A robust sandwich estimator was used in all the models for SEs. The estimated outcome mean changes were compared between the racial and ethnic subgroups. A binary logistic regression was created with non-Hispanic White race as the reference category to address the odds ratio (OR) for being a dropout and for reaching a 12-week pain MCID, adjusting for age, gender, BMI, therapy area, education level, and employment status. A significance level of 0.05 was considered statistically significant. LGCA was coded using R (version 1.4.1717; R Foundation for Statistical Computing), and all other analyses were performed using SPSS (version 17.0; SPSS Inc).

Results

A total of 9550 participants were enrolled, with 6949 (72.8%) patients having completed the program. The study flow diagram is presented in Figure 2.

Figure 2. Flowchart of the study following the CONSORT (Consolidated Standards of Reporting Trials) guidelines.



Baseline Characteristics

The patients' baseline demographics for the entire cohort and for the different race and ethnic groups are presented in Table 1.

On average, participants had 49.4 (SD 12.9) years, a BMI of 29.2 (SD 6.7), and a pain score of 4.9 (SD 2.0). The cohort comprised 58.5% (5589/9550) women, 41.1% (3929/9550) men, 0.3% (24/9550) nonbinary patients, and 0.1% (8/9550) preferred not to answer. Therapy area distribution was similar to the prevalence reported for each MSK pain condition according to the United States Bone and Joint Initiative [43]. The self-reported race and ethnicity groups followed the report for the US population based on the 2020 US census [44] (Multimedia Appendix 1 [44-46], Figure S1).

At baseline, Black and non-Hispanic White patients had a significantly higher mean age than the other patients (P<.001; Table 1). The Black patient group included patients with higher BMI levels (P<.001), a higher proportion of women (P<.001), and those with low back pain (P<.001; Table 1). Asian patients were, on average, the youngest (P<.001) and reported the lowest average BMI score (P<.001; Table 1). Asian patients presented a higher proportion of individuals with higher education, whereas Black and Hispanic patients reported the highest proportion of patients with high school or lower education levels (P<.001; Figure 3).

A larger proportion of full-time employed patients was observed within the Asian and Hispanic patient groups than in the other groups (P<.001).

Regarding clinical outcomes, Black and Hispanic patients reported the highest level of pain, surgical intent, work productivity, and activities of daily living impairment at baseline (P<.001; Table 1). Asian patients reported lower anxiety and depression burdens (P<.001; Table 1).

Comparing completers (n=6949) with noncompleters (n=2601), no differences were observed between the proportions of the different race and ethnic groups (P=.26). Completers were older (50.0, SD 12.7 vs 47.8, SD 13.4, P<.001), with more patients reporting knee and shoulder pain and fewer patients reporting low back pain (P<.001). In addition, completers had a higher proportion of patients with a postgraduate education (P<.001). No differences were observed in employment status.

No clinically relevant differences were observed in clinical outcomes at baseline, despite the statistical differences found between the groups (an effect of the large sample size). For example, pain levels were 5.0 (SD 2.0) in noncompleters and 4.9 (SD 2.0) in completers (Multimedia Appendix 1, Table S1).

Table 1. Baseline characteristics for each racial and ethnic group and for the entire cohort.

Characteristic	Asian (n=910)	Black (n=1025)	Hispanic (n=913)	Non-Hispanic White (n=6240)	Other (462)	*P* value	Entire cohort
Age (years), mean (SD)	44.4 (11.2)	50.4 (12.4)	45.8 (11.4)	50.7 (13.2)	46.1 (12.5)	<.001	49.4 (12.9)
Gender, n (%)						<.001	
Woman	485 (53.3)	713 (69.6)	497 (54.5)	3642 (58.4)	251 (54.3)		5589 (58.5)
Man	424 (46.6)	311 (30.3)	412 (45.1)	2576 (41.3)	206 (44.6)		3929 (41.1)
Nonbinary	1 (0.1)	0 (0)	3 (0.3)	19 (0.3)	1 (0.2)		24 (0.3)
Prefers not to answer	0 (0)	0 (0)	1 (0.0)	3 (0.0)	4 (0.9)		8 (0.1)
BMI, mean (SD)	25.3 (4.4)	31.7 (6.9)	29.8 (6.4)	29.4 (6.7)	28.3 (6.2)		29.2 (6.7)
BMI category n (%)						<.001	
Class III obese	5 (0.5)	135 (13.2)	66 (7.2)	476 (7.6)	24 (5.2)		706 (7.4)
Obese	110 (12.1)	422 (14.6)	299 (32.7)	1932 (31.0)	120 (26.0)		2883 (30.2)
Overweight	317 (34.8)	318 (31.0)	350 (38.3)	2109 (33.8)	168 (36.4)		3262 (34.2)
Healthy	460 (50.5)	144 (14.0)	192 (21.0)	1676 (26.9)	142 (30.7)		2614 (27.4)
Underweight	18 (2.0)	6 (0.6)	6 (0.7)	47 (0.8)	8 (1.7)		85 (0.9)
Therapy area, n (%)						<.001	
Ankle	36 (4.0)	47 (4.6)	37 (4.1)	216 (3.5)	16 (3.5)		*352 (3.7)*
Elbow	17 (1.9)	10 (1.0)	12 (1.3)	140 (2.2)	12 (2.6)		191 (2.0)
Hip	44 (4.8)	84 (8.2)	72 (7.9)	669 (10.7)	43 (9.3)		817 (8.6)
Knee	105 (11.5)	176 (17.2)	115 (12.6)	813 (13.0)	66 (14.3)		1275 (13.4)
Low back	349 (38.4)	505 (49.3)	394 (43.2)	2735 (43.8)	189 (40.9)		4097 (42.9)
Neck	116 (12.7)	56 (5.5)	85 (9.3)	577 (9.2)	48 (10.4)		882 (9.2)
Shoulder	195 (21.4)	121 (11.8)	146 (16.0)	896 (14.4)	73 (15.8)		1431 (15.0)
Wrist and hand	48 (5.3)	26 (2.5)	52 (5.7)	194 (3.1)	15 (3.2)		335 (3.5)
Employment status, n (%)						<.001	
Employed full time	822 (90.3)	804 (78.4)	777 (85.1)	4886 (78.3)	364 (78.8)		7653 (80.1)
Employed part-time	22 (2.4)	36 (3.5)	37 (4.1)	317 (5.1)	15 (3.2)		427 (4.5)
Not employed	21 (2.3)	37 (3.6)	40 (4.4)	303 (4.9)	13 (2.8)		414 (4.3)
Prefers not to answer	21 (2.3)	12 (1.2)	11 (1.2)	65 (1.0)	30 (6.5)		139 (1.5)
Retired	15 (1.6)	117 (11.4)	32 (3.5)	604 (9.7)	28 (6.1)		796 (8.3)
Seeking opportunities	9 (1.0)	5 (0.6)	8 (0.9)	38 (0.6)	6 (1.3)		66 (0.7)
Student	1 (0.0)	14 (1.4)	8 (0.9)	27 (0.4)	6 (1.3)		55 (0.6)
Education level, n (%)						<.001	
Some elementary or middle school	0 (0)	1 (0.1)	2 (0.2)	3 (0.0)	0 (0)		*6 (0.1)*
Some high school	2 (0.2)	10 (1.0)	13 (1.4)	35 (0.6)	2 (0.4)		62 (0.6)
High school graduate or GED[a] (includes technical or vocational training)	25 (2.7)	138 (13.5)	160 (17.5)	630 (10.1)	41 (8.9)		994 (10.4)
Some college (some community college, associate degree)	79 (8.7)	398 (38.8)	279 (30.6)	1731 (27.7)	100 (21.6)		2587 (27.1)
4-year college degree or bachelor's degree	422 (46.4)	260 (25.4)	270 (29.6)	2161 (34.6)	129 (27.9)		3242 (33.9)
Some postgraduate or professional schooling, no postgraduate degree	29 (3.2)	33 (3.2)	29 (3.2)	223 (3.6)	15 (3.2)		329 (3.4)

Characteristic	Asian (n=910)	Black (n=1025)	Hispanic (n=913)	Non-Hispanic White (n=6240)	Other (462)	*P* value	Entire cohort
Postgraduate or professional degree	346 (38.0)	179 (17.5)	144 (15.8)	1416 (22.7)	107 (23.2)		2192 (23.0)
Prefers not to answer	7 (0.8)	6 (0.6)	15 (1.6)	41 (0.7)	68 (14.7)		137 (1.4)
Clinical outcomes, mean (SD)						<.001	
Pain level	4.7 (2.1)	5.6 (2.1)	5.3 (2.0)	4.8 (2.0)	4.8 (2.0)		*4.9 (2.0)*
Surgery intent >0	*19.0 (20.1)*	*29.3 (27.8)*	*26.3 (24.9)*	*24.4 (24.6)*	*21.9 (22.7)*		*24.5 (24.7)*
Surgery intent	7.4 (15.6)	13.1 (23.6)	10.8 (20.5)	10.4 (20.0)	8.2 (17.4)8		*10.3 (20.1)*
GAD-7[b]≥5	*8.1 (3.5)*	*9.3 (4.2)*	*9.9 (4.7)*	*8.8 (4.0)*	*9.5 (4.3)*		*9.0 (4.1)*
GAD-7	2.7 (4.0)	3.0 (4.6)	4.0 (5.3)	3.2 (4.5)	3.3 (4.8)		3.2 (4.5)
PHQ-9[c]≥5	8.2 (3.6)	9.5 (4.4)	10.0 (5.0)	9.5 (4.3)	10.0 (4.9)		9.5 (4.4)
PHQ-9	1.8 (3.5)	2.7 (4.6)	2.8 (3.5)	2.5 (4.5)	2.8 (4.9)		2.5 (4.5)
WPAI[d] overall>0	27.6 (18.3)	35.6 (21.2)	33.6 (22.5)	29.1 (19.2)	29.7 (19.6)		30.1 (19.8)
WPAI overall	16.0 (19.5)	20.3 (23.8)	19.5 (23.9)	17.3 (20.6)	18.4 (21.2)		17.7 (21.2)
WPAI work >0	26.5 (17.3)	34.3 (20.3)	32.3 (21.6)	28.2 (18.2)	28.5 (18.5)		29.0 (18.8)
WPAI work	15.1 (18.5)	19.1 (22.8)	18.4 (22.8)	16.4 (19.6)	17.4 (20.1)		16.8 (20.2)
WPAI time >0	19.1 (23.7)	37.1 (36.2)	28.6 (33.6)	23.1 (27.7)	29.4 (33.9)		25.5 (30.0)
WPAI time	2.1 (9.9)	4.9 (18.2)	3.8 (15.7)	2.3 (11.2)	4.7 (17.3)		2.8 (12.8)
WPAI activity >0	33.0 (21.7)	42.2 (24.1)	39.7 (24.1)	37.3 (22.2)	37.8 (22.7)		37.6 (22.6)
WPAI activity	23.4 (23.6)	30.1 (27.9)	28.8 (27.1)	30.0 (24.8)	29.8 (25.4)		29.3 (25.3)

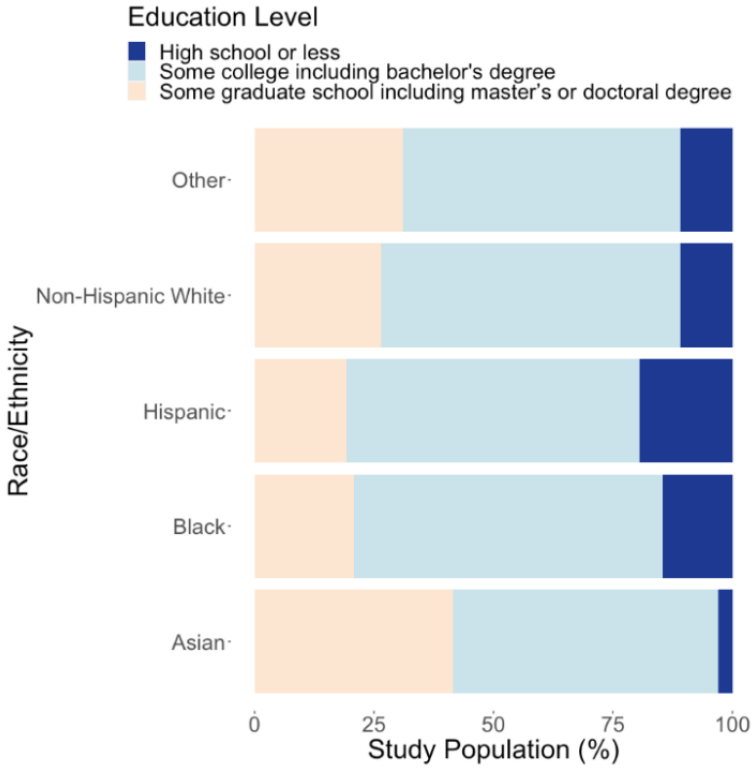
[a]GED: General Educational Development.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PHQ-9: Patient Health Questionnaire 9-item.

[d]WPAI: Work Productivity and Activity Impairment questionnaire.

Figure 3. Distribution of different education levels across the different race and ethnic groups.



Engagement Outcomes

The overall completion rate was 72.8% (6949/9550). When stratifying dropouts by racial and ethnic groups, 20% (184/910) Asian, 23% (231/1025) Black, 25% (232/913) Hispanic, 20% (1224/6240) non-Hispanic White, and 20% (94/462) other racial and ethnic groups patients dropped out by the end of the program. The OR for being a dropout was estimated having non-Hispanic White patients as reference, Black patients: 1.14, 95% CI 0.97-1.34; Asian patients: 1.03, 95% CI 0.86-1.23; Hispanic patients: 1.19 95% CI 1.01-1.40; other patients: 0.95, 95% CI 0.74-1.21. Both Hispanic and Black patients seemed more likely to drop out than non-Hispanic White patients, although only Hispanic patients' OR reached statistical significance (P=.04).

The studied covariates influenced the obtained OR, with men (P=.006), younger patients (P<.001), patients with higher BMI scores (P<.001), less educated (P<.001), and those with spine conditions (P=.04) being more likely to drop out.

Completers performed an average of 30.1 (SD 20.0) sessions, comprising an average of 355.4 (SD 239.6) minutes of training time at an average of 2.8 (SD 1.1) sessions per week (Table 2). The mean number of education articles read was 2.7 (SD 1.1), and the mean number of interactions with PT was 16.0 (SD 14.0), whereas mean satisfaction score was 9.0 (SD 1.5; Table 2).

Across the different racial and ethnic groups, Black, Hispanic, and other patients participated in significantly fewer total sessions (P<.001, P<.001, and P=.001, respectively), had less training time (P<.001, P<.001, and P=.001, respectively), and lower average number of sessions per week (P<.001, P<.001, and P=.003, respectively) when compared with non-Hispanic White patients. Black and non-Hispanic White patients read more articles (P<.001). Black and Hispanic patients were more satisfied with their treatment results (P values ranging from <0.001 to 0.009; Table 2). Black patients had a significantly lower mean number of interactions with PT than non-Hispanic White patients (P<.001).

Table 2. Twelve-week program engagement data across the racial and ethnic groups following an intention-to-treat and per-protocol analysis. All values are mean (SD) values.

Analysis	Asian		Black		Hispanic		Non-Hispanic White		Other		*P* value		Entire cohort	
	ITT[a]	PP[b]	ITT	PP	ITT	PP	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Total number of sessions	24.7 (20.3)	30.5 (20.2)	21.6 (19.8)	26.9 (20.4)	20.5 (17.3)	26.2 (17.4)	25.4 (20.3)	31.4 (20.2)	22.1 (18.8)	27.1 (19.0)	<.001	<.001	24.3 (20.0)	30.1 (20.0)
Total time on sessions	288.4 (236.5)	356.7 (233.7)	255.8 (238.6)	320.1 (247.1)	253.4 (220.3)	325.2 (221.2)	297.2 (242.6)	368.3 (242.0)	254.9 (215.6)	315.7 (220.1)	<.001	<.001	285.6 (238.9)	355.4 (239.6)
Number of sessions per week	2.6 (1.1)	2.8 (1.1)	2.4 (1.1)	2.6 (1.1)	2.4 (0.9)	2.5 (0.9)	2.7 (1.1)	2.9 (1.1)	2.5 (1.0)	2.6 (1.0)	<.001	<.001	2.6 (1.1)	2.8 (1.1)
Total articles read	1.6 (3.1)	1.9 (3.4)	2.4 (4.6)	2.8 (5.1)	2.1 (4.1)	2.4 (4.6)	2.5 (4.6)	2.8 (5.0)	2.1 (3.7)	2.3 (3.9)	<.001	<.001	2.3 (4.4)	2.7 (1.1)
Total interactions with PT[c]	12.9 (12.3)	15.1 (13.1)	11.2 (12.0)	13.3 (12.9)	12.6 (12.3)	15.2 (13.1)	14.3 (13.5)	16.7 (14.3)	13.3 (13.6)	15.1 (14.5)	<.001	<.001	13.1 (13.1)	16.0 (14.0)
Overall satisfaction	8.8 (1.4)	8.9 (1.4)	9.3 (1.1)	9.3 (1.1)	9.3 (1.3)	9.3 (1.3)	8.9 (1.5)	8.9 (1.5)	8.8 (1.7)	8.8 (1.7)	<.001	<.001	9.0 (1.5)	9.0 (1.5)

[a]ITT: intention-to-treat analysis.

[b]PP: per-protocol analysis.

[c]PT: physical therapist.

Clinical Outcomes at Program End (12 Weeks)

Clinical outcomes at end of the program for each race and ethnicity were examined following both an intention-to-treat and per-protocol analysis, as presented in Table 3 (for outcomes unfiltered at baseline please see Multimedia Appendix 1, Table S2). The LGCA models for both intention-to-treat and per-protocol are presented in Multimedia Appendix 1 Tables S3 and S4, respectively. Both models presented good fit, as shown in Multimedia Appendix 1, Table S5. Both analyses provided very similar results, probably because of the combination of large sample sizes and high completion rates. The presentation of the results will focus on per-protocol analysis, as it is more truly reflective of the impact of the program on clinical outcomes.

Table 3. Baseline and 12-week estimated outcome metrics following an ITT and PP analysis for each of the racial and ethnic groups (outcomes filtered at baseline as explained in the table)[a].

Outcome and time	Asian		Black		Hispanic		Non-Hispanic White		Other		Entire cohort	
	ITT[b]	PP[c]	ITT	PP	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Pain level, mean (95% CI)												
Baseline	4.6 (4.5-4.8)	4.6 (4.5-4.8)	5.6 (5.4-5.7)	5.5 (85.4-5.7)	5.3 (5.1-5.4)	5.3 (5.1-5.5)	4.8 (4.7-4.8)	4.7 (4.6-4.7)	4.8 (4.6-5.0)	4.7 (4.5-4.9)	4.9 (4.8-4.9)	4.8 (4.8-4.9)
12 weeks	2.6 (2.4-2.8)	2.6 (2.4-2.8)	3.2 (3.0-3.5)	3.2 (2.9-3.5)	2.7 (2.5-3.0)	2.7 (2.4-2.9)	2.8 (2.8-2.9)	2.8 (2.7-2.9)	2.9 (2.6-3.2)	2.8 (2.5-3.1)	2.9 (2.8-2.9)	2.8 (2.7-2.9)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Mean change, OR (95% CI)	2.02 (1.80-2.25)	2.00 (1.77-2.24)	2.35 (2.10-2.61)	2.35 (2.1-2.6)	2.55 (2.30-2.81)	2.63 (2.37-2.88)	1.91 (1.82-1.99)	1.88 (1.79-1.97)	1.88 (1.56-2.21)	1.90 (1.58-2.22)	2.0 (1.9-2.1)	2.0 (1.9-2.1)
Surgery intent>0, mean (95% CI)												
Baseline	18.6 (16.6-20.7)	17.4 (15.1-19.6)	28.8 (26.3-31.3)	27.7 (24.8-30.6)	25.9 (23.4-28.4)	25.6 (22.6-28.7)	24.1 (23.1-25.0)	22.8 (21.8-23.9)	21.5 (18.1-25.0)	21.1 (17.2-25.0)	24.2 (23.4-24.9)	23.1 (22.2-24.0)
12 weeks	9.1 (6.0-12.2)	8.6 (5.4-11.8)	15.0 (11.0-19.0)	14.1 (10.0-18.2)	11.8 (8.5-15.0)	11.1 (7.8-14.4)	13.5 (12.1-14.9)	12.3 (10.9-13.7)	10.0 (5.6-14.4)	9.9 (5.2-14.6)	13.0 (11.9-14.2)	12.0 (10.9-13.2)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Mean change	9.54 (6.14-12.93)	8.79 (5.28-12.29)	13.85 (9.80-17.91)	13.62 (9.48-17.75)	14.16 (10.85-17.46)	14.46 (11.14-17.98)	10.54 (9.22-11.85)	10.51 (9.20-11.83)	11.56 (7.46-15.65)	11.18 (6.89-15.47)	11.1 (10.0-12.3)	11.1 (9.9-12.2)
GAD-7[d] ≥5, mean (95% CI)												
Baseline	8.1 (7.6-8.5)	8.0 (7.5-8.5)	9.2 (8.8-9.7)	9.0 (8.4-9.6)	9.9 (9.4-10.4)	9.8 (9.1-10.4)	8.8 (8.6-9.0)	8.6 (8.4-8.8)	9.5 (8.8-10.2)	9.1 (8.3-9.9)	8.9 (8.8-9.1)	8.7 (8.5-8.9)
12 weeks	3.6 (2.9-4.3)	3.7 (2.9-4.4)	4.3 (3.5-5.1)	4.1 (3.3-4.9)	4.9 (3.8-6.1)	4.9 (3.7-6.1)	4.9 (4.6-5.3)	4.8 (4.4-5.2)	5.2 (3.6-6.8)	5.0 (3.3-6.7)	4.8 (4.5-5.1)	4.7 (4.4-5.0)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Mean change	4.44 (3.62-5.26)	4.37 (3.51-5.22)	4.92 (4.18-5.66)	4.91 (4.17-6.65)	4.93 (3.78-6.08)	4.86 (3.68-6.04)	3.87 (3.55-4.20)	3.77 (3.44-4.10)	4.30 (2.65-5.95)	4.07 (2.33-5.81)	4.1 (3.8-4.4)	4.0 (3.7-4.3)
PHQ-9[e] ≥5, mean (95% CI)												
Baseline	8.2 (7.7-8.8)	8.0 (7.4-8.7)	9.5 (8.9-10.0)	9.0 (8.4-3.2)	10.0 (9.4-10.7)	9.6 (8.8-10.4)	9.5 (9.2-9.7)	9.2 (9.0-9.5)	10.0 (9.1-10.9)	9.5 (8.4-10.5)	9.5 (9.3-9.7)	9.1 (8.9-9.4)
12 weeks	3.8 (2.7-4.8)	3.6 (2.5-4.7)	4.7 (3.5-5.9)	4.4 (3.2-5.5)	6.4 (4.9-7.9)	6.2 (4.7-7.8)	5.2 (4.8-5.7)	5.0 (4.5-5.4)	6.6 (4.7-8.4)	6.2 (4.3-8.1)	5.2 (4.9-5.6)	5.0 (4.6-5.4)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001	<.001
Mean change	4.49 (3.47-5.52)	4.41 (3.36-5.46)	4.80 (3.63-5.96)	4.66 (3.48-5.83)	3.68 (2.19-5.16)	3.38 (1.86-4.89)	4.30 (3.86-4.74)	4.24 (3.80-4.68)	3.41 (1.54-5.29)	3.29 (1.33-5.26)	4.3 (4.6-3.9)	4.2 (3.8-4.5)
WPAI overall work impairment[f] >0, mean (95% CI)												
Baseline	27.5 (25.8-29.1)	27.1 (25.2-29.0)	35.6 (33.6-37.6)	35.0 (32.7-37.3)	33.5 (31.4-35.6)	33.1 (30.5-35.6)	29.0 (28.3-29.7)	28.3 (27.5-29.2)	29.4 (26.8-32.0)	28.9 (26.0-31.8)	29.9 (29.4-30.5)	29.4 (28.7-30.0)
12 weeks	13.3 (10.1-16.5)	12.7 (9.5-15.9)	15.9 (12.3-19.6)	15.9 (12.2-19.6)	18.3 (13.6-23.1)	18.3 (13.5-23.1)	16.0 (14.7-17.4)	15.5 (14.1-16.8)	17.8 (13.3-22.4)	17.8 (13.1-22.5)	16.1 (15.0-17.3)	15.7 (14.5-18.8)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Outcome and time	Asian		Black		Hispanic		Non-Hispanic White		Other		Entire cohort	
	ITT[b]	PP[c]	ITT	PP	ITT	PP	ITT	PP	ITT	PP	ITT	PP
Mean change	14.12 (10.83-17.40)	14.41 (11.08-17.74)	19.64 (15.91-23.37)	19.14 (15.30-22.97)	15.18 (10.25-20.10)	14.80 (9.78-19.83)	12.96 (11.61-14.31)	12.86 (11.49-14.22)	11.55 (6.93-16.17)	11.11 (6.32-15.90)	13.8 (12.6-15.0)	13.7 (12.5-14.9)
WPAI work impairment>0, mean (95% CI)												
Baseline	26.3 (24.7-27.9)	25.9 (24.1-27.7)	34.2 (32.3-36.1)	33.7 (31.5-36.0)	32.1 (30.1-34.1)	31.3 (28.9-33.7)	28.0 (27.4-28.7)	27.3 (26.5-28.1)	28.2 (25.7-30.6)	27.5 (24.8-30.2)	28.8 (28.3-29.4)	28.2 (27.5-28.8)
12 weeks	11.7 (9.0-14.4)	11.1 (8.4-13.8)	14.8 (11.4-18.3)	14.8 (11.2-18.3)	17.2 (12.6-21.7)	17.0 (12.4-21.6)	14.8 (13.6-16.1)	14.3 (13.0-15.5)	16.2 (12.2-20.3)	16.2 (12.0-20.4)	14.8 (13.8-15.9)	14.4 (13.3-15.5)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Mean change	14.57 (11.75-17.38)	14.80 (11.96-17.65)	19.35 (15.74-22.96)	18.97 (15.24-22.70)	14.91 (10.20-19.62)	14.35 (9.55-19.15)	13.21 (11.95-14.46)	13.03 (11.76-14.31)	11.92 87.76-16.08)	11.35 (7.05-15.66)	14.0 (12.9-15.1)	13.8 (12.7-14.9)
WPAI work time missed>0, mean (95% CI)												
Baseline	19.0 (14.0-24.0)	20.2 (14.1-26.4)	37.4 (30.5-44.4)	33.0 (25.2-40.8	28.5 (22.1-35.0)	26.8 (19.9-33.7)	23.0 (20.5-25.4)	21.6 (19.0-24.3)	29.5 (20.9-38.1)	26.8 (17.0-36.5)	25.5 (23.4-27.5)	23.9 (21.7-26.2)
12 weeks	7.7 (0-17.0)	7.1 (0-16.3)	13.2 (4.0-22.4)	13.2 (4.1-22.2)	7.3 (3.0-11.5)	7.4 (2.8-12.0)	8.4 (5.5-11.3)	8.2 (5.3-11.1)	5.6 (0-14.24)	16.2 (12.0-20.4)	8.7 (6.1-11.2)	8.5 (6.1-11.0)
P value	0.04	0.03	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.001	<.001	<.001
Mean change	11.27 (0.36-22.17)	13.08 (1.55-24.61)	24.20 (14.83-33.58)	19.81 (10.84-28.78)	21.15 (15.02-27.48)	19.42 (13.04-25.81)	14.55 (11.06-18.03)	13.44 (9.88-17.00)	23.92 (11.47-36.37)	21.70 (8.51-34.88)	16.8 (19.8-13.8)	15.4 (12.5-18.4)
WPAI activity impairment>0, mean (95% CI)												
Baseline	32.8 (31.1-34.4)	33.2 (31.3-35.1)	42.0 (40.3-43.8)	41.9 (39.9-43.9)	39.5 (37.7-41.3)	38.4 (36.2-40.5)	37.1 (36.5-37.7)	36.2 (35.5-37.0)	37.7 (35.4-40.1)	37.6 (34.9-40.4)	37.4 (36.9-38.0)	36.8 (36.2-37.4)
12 weeks	15.7 (13.3 - 18.2)	15.7 (13.1-18.2)	21.6 (18.3-24.8)	21.6 (18.3-24.9)	18.5 (15.4-21.5)	17.9 (14.8-21.0)	20.2 (19.2-21.3)	19.6 (18.5-20.6)	23.2 (19.2-27.5)	22.3 (18.3-26.4)	20.1 (19.2-20.9)	19.5 (18.6-20.4)
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Mean changes	17.05 (14.39-19.71)	17.53 (14.76-20.31)	20.49 (17.13-23.84)	20.34 (16.87-23.80)	21.03 (17.89-24.18)	20.50 (17.27-23.73)	16.85 (15.82-17.88)	16.67 (15.62-17.72)	14.40 (10.1-18.7)	15.26 (10.95-19.57)	17.4 (16.5-18.3)	17.3 (16.4-18.2)

[a]Data represent the mean (95% CI). *P* values represent comparisons between 12-week and baseline means with statistically significant *P* values italicized.

[b]ITT: intention-to-treat analysis.

[c]PP: per-protocol analysis.

[d]GAD-7: Generalized Anxiety Disorder 7-item scale.

[e]PHQ-9: Patient Health 9-item questionnaire.

[f]WPAI: Work Productivity and Activity Impairment Questionnaire.

Pain Scores and Pain MCID

Patients experienced a significant reduction in mean pain scores at 12 weeks compared with baseline across all racial and ethnic groups (*P*<.001 for each analysis; Table 3). Black and Hispanic patients had a significantly larger reduction in mean pain level scores than non-Hispanic White patients (*P*=.001 and *P*<.001, respectively) and those in the other groups (*P*=.03 and *P*=.001, respectively; Multimedia Appendix 1, Table S6). Of note, both Black and Hispanic patients also had a significantly higher mean baseline pain level than the other groups (*P*<.001; Table 3).

When considering the recommended MCID for pain scores, 75.8% (157/207) Hispanic patients had a greater response rate at the 12-week assessment when compared with all other groups (Black patients: 167/246, 66.4%, *P*=.03; non-Hispanic White patients: 1177/1841, 63.9%, *P*<.001; and other patients: 76/126, 60.3%, *P*=.003), with the exception of Asian patients (167/246, 67.9%, *P*=.06)

To evaluate whether race and ethnicity was an independent factor for reaching MCID, logistic regression adjusted for BMI, age, sex, therapy area, education level, and employment status was performed with the non-Hispanic White race as the reference category. Hispanic patients (OR 1.74, 95% CI 1.24-2.45) were more likely to achieve MCID than non-Hispanic White patients (P=.001). The OR for the other race groups did not reach statistical significance. Both men (P=.007) and patients with upper limb pain (P<.001) were more likely to achieve MCID.

Surgery Intent

The mean surgical intent score was significantly reduced overall (11.1, 95% CI 9.9-12.2, P<.001) and within each racial and ethnic group at 12 weeks (Table 3). Hispanic patients reported a higher reduction in the willingness to pursue surgery (14.46, 95% CI 11.14-17.98), which was statistically different from Asian patients (P=.02) and non-Hispanic White patients (P=.03; Multimedia Appendix 1, Table S6), followed by Black patients (13.62, 95% CI 9.48-17.75), which were only statistically different from Asian patients (P=.08).

Mental Health (GAD-7 and PHQ-9)

A significant improvement in both mental health metrics was observed for the overall cohort compared with baseline when filtering for at least mild anxiety and depression at baseline (scores above 5) (GAD-7: 4.0, 95% CI 3.7-4.3, P<.001; and PHQ-9: 4.2, 95% CI 3.8-4.5, P<.001). Reductions were similar across all racial and ethnic groups in both anxiety and depression mean changes, with scores ranging between 3.19 and 4.91. Black patients exhibited the greatest reduction in GAD-7 (4.91, 95% CI 4.17-6.65), which was statistically different from non-Hispanic White patients (3.77, 95% CI 3.44-4.10, P=.005), but not clinically relevant.

Work Productivity

For the overall cohort, there was a significant improvement in all WPAI domains compared with baseline: WPAI overall: 13.7, 95% CI 12.5-14.9, P<.001; WPAI work: 13.8, 95% CI 12.7-14.9, P<.001; WPAI time: 15.4, 95% CI 12.5-18.4, P<.001; WPAI activity: 17.3, 95% CI 16.4-18.2, P<.001. Each racial and ethnic group experienced a significant improvement in the mean WPAI overall, WPAI work, WPAI time, and WPAI activity scores (P<.001; Table 3). Black patients recovered the most from presenteeism (18.97, 95% CI 15.24-22.70), a change statistically different from non-Hispanic White patients (13.03, 95% CI 11.76-14.31, P=.003) and from other patients (11.35, 95% CI 7.05-15.66, P=.008). Both Black (20.34, 95% CI 16.87-23.80) and Hispanic patients (20.50, 95% CI 17.27-23.73) recovered more from activities of daily living impairment than non-Hispanic White patients (16.67, 95% CI 15.62-17.72, P=.046 and P=.03, respectively).

Discussion

Principal Findings

Among the racial and ethnic groups studied, Black patients presented baseline demographic characteristics associated with poorer prognosis (higher prevalence of women [47], older patients [48], and those with higher BMI levels [49]), whereas Asian patients were the youngest and reported the lowest average BMI score. Asian patients presented a higher proportion of individuals with high education levels, whereas Black and Hispanic patients reported the highest proportion of patients with high school or lower education levels.

Overall, completion rates, engagement, and satisfaction levels were high. However, Black patients had a higher OR for dropping out with Hispanic patients showing the same tendency. Black, Hispanic, and other patients engaged less with the program, but both Black and Hispanic patients reported more overall satisfaction with the DCP. Black patients interacted the least with PT but read more articles (alongside non-Hispanic White) than patients from other races and ethnicities.

Regarding the clinical outcomes, significant pain reduction was observed in all racial and ethnic groups. Black and Hispanic patients reported the highest level of pain, surgical intent, work productivity, and impairment in activities of daily living at baseline. However, these same patients also reported the greatest reduction in surgery intention, work productivity, and activities of daily living impairment by program end, when compared with the other racial and ethnic groups. Black and Hispanic patients had a larger reduction in mean pain level scores than non-Hispanic White patients and those from the other groups; however, only Hispanic patients reported significantly greater response rates (157/207, 75.8%).

Comparison With Prior Work

To our knowledge, this is the first study to evaluate racial differences in engagement and outcomes for a completely remote, multimodal, digital care plan for MSK pain. Several reports have shown that people from racial and ethnic minority groups do not access telehealth as often as non-Hispanic White patients [21,27,32,33]. However, in this study, the distribution of different racial and ethnic groups that enrolled in the study followed the proportions in the US population [44], which is a testament to the accessibility of a DCP offered through employers' health plans.

Overall engagement in the program was high, with a high satisfaction rate. Black and Hispanic patients dropped out more frequently than the other groups and had lower metrics for engagement. However, these 2 groups also had the highest satisfaction scores. Different combinations of factors might explain the lower engagement of Black and Hispanic patients with DCP. Aggravated baseline outcomes may be associated with poorer adherence [50,51]. Among demographic characteristics, high BMI scores (as observed in Black patients) have been associated with lower treatment adherence rates [49,52]. The higher proportion of patients with lower educational levels within the Black and Hispanic groups may partially contribute to lower engagement rates. However, this may not be causal, as it is well known that patients with poor digital literacy have a harder time accessing telehealth services [21-23], and that individuals with lower education levels have lower digital literacy [53]. Given that racial and ethnic enrollment in our study was proportional to the US population, it would appear that employer-based health care plans have helped remove access barriers to digital rehabilitation. Nevertheless, our

findings suggest that society at large should focus on tailored engagement strategies in these groups, as program completers tend to experience better outcomes than dropouts.

Significant improvements in pain were observed at the completion of the program across all different racial and ethnic groups. However, it is known that pain is not equally experienced among different races and ethnicities [24,25,30,31]. People from racial and ethnic minority groups have been reported to experience higher levels of pain and disability [28,29]. This was observed in this study, with both Black and Hispanic patients having significantly higher baseline pain scores.

In addition, people from racial and ethnic minority groups have been shown to have worse outcomes than non-Hispanic White patients [25,28,33]. However, this was not observed in this study. Both Black and Hispanic patients had significantly larger improvements in pain at the completion of the study, with Hispanic patients reporting higher odds of reaching the 30% pain MCID independent of age, BMI, therapy area, education level, sex, and employment status when compared with non-Hispanic White patients. This trend was similar to that for work productivity improvement. All patients showed significant improvement at the completion of the program in all WPAI subdomains, with Black and Hispanic patients having significantly larger improvements. It is important to note that both groups had higher baseline pain and WPAI scores, and thus, more room to improve. Despite this, the results are still striking and advocate for digital therapy for MSK pain in these populations.

Black and Hispanic patients also had significantly higher baseline surgical intentions, which was not surprising given their higher pain scores. To our knowledge, no study has investigated racial or ethnic differences in surgical intent in a physical rehabilitation setting, which makes comparisons difficult.

It is well established that MSK pain is associated with comorbid psychiatric illnesses, specifically depression and anxiety [54]. In this study, all patients showed improved mental health metrics for depression and anxiety, which were not significantly different when stratified by race and ethnicity. This finding supports the notion that all groups benefited similarly from the program in terms of mental health improvement.

Limitations and Strengths

This study has several limitations, the most relevant being the lack of a control group, which means that we cannot establish the program's causal effect on pain or other clinical outcome improvements. Nevertheless, the large sample size and applied statistical analysis allowed not only to compare clinical status in a before and after scenario but also to compare the trajectories of distinct groups of patients, which was the main intent of this study. In addition, the fact that all patients had chronic MSK conditions provides a more homogeneous sample, where the natural history of the condition tends not to be as favorable as in cohorts of patients, including acute MSK pain.

Our study participants may not be representative of the general adult population, as the study only included beneficiaries of specific benefits provided by their employers or covered by health plans offering the service, and who opted into a digital MSK program, which limits their applicability to clinical settings with higher proportions of uninsured, elder adults, or patients who are work-disabled.

This study also does not control for all domains known as social determinants of health (eg, income), which can influence both program use and health outcomes, and are known to disproportionately affect different racial and ethnic groups [24-27]. Long-term follow-up was also not available to ascertain the benefits of the program at later time points and to determine whether any racial differences remained or dissipated.

Further prospective controlled studies are warranted to better characterize the effects of race and ethnicity on digital therapy outcomes, namely, controlling for social determinants of health.

Despite these limitations, the results provide evidence of program applicability in a real-world setting with a large sample size from a wide geographic representation (50 states and the District of Columbia in the United States), with a wide diversity of job types (eg, nurses, manual laborers, and office workers). Therefore, this cohort allows for a diverse population study, with large subgroup sample sizes enabling comparisons, which to the best of our knowledge, have not been reported before. Another strength is the DCP itself, which is a multimodal approach that includes exercises using real-time biofeedback, regular communication with the same PT, and a digital format, all of which favor accessibility and maximize engagement. An additional strength of this study is the use of validated outcome metrics for both physical and psychological outcomes, thereby permitting translational application and generalizability to other populations.

Conclusions

This study is the first to evaluate racial differences in a completely remote, multimodal, DCP for MSK pain. The study population followed the proportions in the US population. All racial and ethnic groups experienced significant improvements in pain as well as high satisfaction rates at program completion. Black and Hispanic patients had significantly higher baseline outcome scores, lower engagement metrics, and higher dropout rates, but they also had higher satisfaction rates and improvements in those outcomes. Hispanic patients reported the higher response rate to pain. This study supports the use of DCPs to improve accessibility, while reinforcing the need to improve engagement strategies for Black and Hispanic patients.

Acknowledgments

The authors acknowledge the team of physical therapists responsible for managing the participants. The authors also acknowledge the contributions of João Tiago Silva and Guilherme Freches in data validation (all employees of Sword Health). Critical revision

of the manuscript for important intellectual content was done by all authors. All authors were involved with the final approval of the version. The study sponsor, Sword Health, was involved in the study design, data collection, and interpretation and writing of the manuscript. The data sets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors made a significant contribution to the work reported as follows: FDC, JL, and SPC were responsible for the study concept and design, MM acquired the data, RGM performed the statistical analysis, JS, F Costa, AA, DJ, MM, and SPC interpreted the data; JS and F Costa were responsible for drafting the work; and VB was responsible for funding.

Conflicts of Interest

F Costa, DJ, AA, MM, FDC, and VY are employees of Sword Health, the sponsor of this study. FDC, VY, and VB also hold equity in Sword Health, and VB is the CEO of the same company. RGM is an independent scientific consultant responsible for the statistical analysis and received consultant fees from Sword Health. JS, SPC, and JL are independent scientific and clinical consultants who received adviser honorarium from Sword Health.

Multimedia Appendix 1
Supplementary tables and figure.
[PDF File (Adobe PDF File), 573 KB - jmir_v24i10e41306_app1.pdf]

References

1. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2021 Dec 19;396(10267):2006-2017 [FREE Full text] [doi: 10.1016/S0140-6736(20)32340-0] [Medline: 33275908]
2. El-Tallawy SN, Nalamasu R, Salem GI, LeQuang JA, Pergolizzi JV, Christo PJ. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. Pain Ther 2021 Jun;10(1):181-209 [FREE Full text] [doi: 10.1007/s40122-021-00235-2] [Medline: 33575952]
3. Bonanni R, Cariati I, Tancredi V, Iundusi R, Gasbarra E, Tarantino U. Chronic pain in musculoskeletal diseases: do you know your enemy? J Clin Med 2022 May 06;11(9):2609 [FREE Full text] [doi: 10.3390/jcm11092609] [Medline: 35566735]
4. George SZ, Fritz JM, Silfies SP, Schneider MJ, Beneciuk JM, Lentz TA, et al. Interventions for the management of acute and chronic low back pain: revision 2021. J Orthop Sports Phys Ther 2021 Nov;51(11):CPG1-CP60. [doi: 10.2519/jospt.2021.0304] [Medline: 34719942]
5. Gianola S, Bargeri S, Del Castillo G, Corbetta D, Turolla A, Andreano A, et al. Effectiveness of treatments for acute and subacute mechanical non-specific low back pain: a systematic review with network meta-analysis. Br J Sports Med 2022 Jan;56(1):41-50 [FREE Full text] [doi: 10.1136/bjsports-2020-103596] [Medline: 33849907]
6. Stochkendahl MJ, Kjaer P, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, et al. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. Eur Spine J 2018 Jan;27(1):60-75. [doi: 10.1007/s00586-017-5099-2] [Medline: 28429142]
7. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. Musculoskelet Sci Pract 2020 Aug;48:102193 [FREE Full text] [doi: 10.1016/j.msksp.2020.102193] [Medline: 32560876]
8. Havran MA, Bidelspach DE. Virtual physical therapy and telerehabilitation. Phys Med Rehabil Clin N Am 2021 May;32(2):419-428. [doi: 10.1016/j.pmr.2020.12.005] [Medline: 33814066]
9. Seron P, Oliveros MJ, Gutierrez-Arias R, Fuentes-Aspe R, Torres-Castro RC, Merino-Osorio C, et al. Effectiveness of telerehabilitation in physical therapy: a rapid overview. Phys Ther 2021 Jun 01;101(6):pzab053 [FREE Full text] [doi: 10.1093/ptj/pzab053] [Medline: 33561280]
10. Dantas LO, Barreto RP, Ferreira CH. Digital physical therapy in the COVID-19 pandemic. Braz J Phys Ther 2020;24(5):381-383 [FREE Full text] [doi: 10.1016/j.bjpt.2020.04.006] [Medline: 32387004]
11. Turolla A, Rossettini G, Viceconti A, Palese A, Geri T. Musculoskeletal physical therapy during the COVID-19 pandemic: is telerehabilitation the answer? Phys Ther 2020 Aug 12;100(8):1260-1264 [FREE Full text] [doi: 10.1093/ptj/pzaa093] [Medline: 32386218]
12. Grona SL, Bath B, Busch A, Rotter T, Trask C, Harrison E. Use of videoconferencing for physical therapy in people with musculoskeletal conditions: a systematic review. J Telemed Telecare 2018 Jun;24(5):341-355. [doi: 10.1177/1357633X17700781] [Medline: 28403669]
13. Levy CE, Silverman E, Jia H, Geiss M, Omura D. Effects of physical therapy delivery via home video telerehabilitation on functional and health-related quality of life outcomes. J Rehabil Res Dev 2015;52(3):361-370 [FREE Full text] [doi: 10.1682/JRRD.2014.10.0239] [Medline: 26230650]

14. Horsley S, Schock G, Grona SL, Montieth K, Mowat B, Stasiuk K, et al. Use of real-time videoconferencing to deliver physical therapy services: a scoping review of published and emerging evidence. J Telemed Telecare 2020 Dec;26(10):581-589. [doi: 10.1177/1357633X19854647] [Medline: 31213166]
15. Nelson M, Bourke M, Crossley K, Russell T. Telerehabilitation is non-inferior to usual care following total hip replacement - a randomized controlled non-inferiority trial. Physiotherapy 2020 Jun;107:19-27. [doi: 10.1016/j.physio.2019.06.006] [Medline: 32026820]
16. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120. [doi: 10.2106/JBJS.I.01375] [Medline: 21248209]
17. Correia FD, Molinos M, Luís S, Carvalho D, Carvalho C, Costa P, et al. Digitally assisted versus conventional home-based rehabilitation after arthroscopic rotator cuff repair: a randomized controlled trial. Am J Phys Med Rehabil 2022 Mar 01;101(3):237-249 [FREE Full text] [doi: 10.1097/PHM.0000000000001780] [Medline: 33935152]
18. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]
19. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299 [FREE Full text] [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
20. Fatoye F, Gebrye T, Fatoye C, Mbada CE, Olaoye MI, Odole AC, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375 [FREE Full text] [doi: 10.2196/15375] [Medline: 32357128]
21. Dixit N, Van Sebille Y, Crawford GB, Ginex PK, Ortega PF, Chan RJ. Disparities in telehealth use: how should the supportive care community respond? Support Care Cancer 2022 Feb;30(2):1007-1010 [FREE Full text] [doi: 10.1007/s00520-021-06629-4] [Medline: 34668075]
22. Lattimore CM, Kane WJ, Fleming 2nd MA, Martin AN, Mehaffey JH, Smolkin ME, et al. Disparities in telemedicine utilization among surgical patients during COVID-19. PLoS One 2021 Oct 8;16(10):e0258452 [FREE Full text] [doi: 10.1371/journal.pone.0258452] [Medline: 34624059]
23. 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]
24. Bond J, Julion WA, Reed M. Racial discrimination and race-based biases on orthopedic-related outcomes: an integrative review. Orthop Nurs 2022;41(2):103-115. [doi: 10.1097/NOR.0000000000000830] [Medline: 35358128]
25. Bove AM, Hausmann LR, Piva SR, Brach JS, Lewis A, Fitzgerald GK. Race differences in postacute physical therapy utilization and patient-reported function after total knee arthroplasty. Arthritis Care Res (Hoboken) 2022 Jan;74(1):79-88. [doi: 10.1002/acr.24792] [Medline: 34553507]
26. Larsen PD. Racism in health care. Rehabil Nurs 2021;46(2):63-64. [doi: 10.1097/RNJ.0000000000000314] [Medline: 33646726]
27. Wegermann K, Wilder JM, Parish A, Niedzwiecki D, Gellad ZF, Muir AJ, et al. Racial and socioeconomic disparities in utilization of telehealth in patients with liver disease during COVID-19. Dig Dis Sci 2022 Jan;67(1):93-99 [FREE Full text] [doi: 10.1007/s10620-021-06842-5] [Medline: 33507442]
28. Hooten WM, Knight-Brown M, Townsend CO, Laures HJ. Clinical outcomes of multidisciplinary pain rehabilitation among African American compared with Caucasian patients with chronic pain. Pain Med 2012 Nov;13(11):1499-1508. [doi: 10.1111/j.1526-4637.2012.01489.x] [Medline: 22994358]
29. Jarvik JG, Comstock BA, Heagerty PJ, Turner JA, Sullivan SD, Shi X, et al. Back pain in seniors: the Back pain Outcomes using Longitudinal Data (BOLD) cohort baseline data. BMC Musculoskelet Disord 2014 Apr 23;15:134 [FREE Full text] [doi: 10.1186/1471-2474-15-134] [Medline: 24755158]
30. Drwecki BB, Moore CF, Ward SE, Prkachin KM. Reducing racial disparities in pain treatment: the role of empathy and perspective-taking. Pain 2011 May;152(5):1001-1006. [doi: 10.1016/j.pain.2010.12.005] [Medline: 21277087]
31. Siddiqui N, Urman RD. Opioid use disorder and racial/ethnic health disparities: prevention and management. Curr Pain Headache Rep 2022 Feb;26(2):129-137. [doi: 10.1007/s11916-022-01010-4] [Medline: 35179723]
32. Weber E, Miller SJ, Astha V, Janevic T, Benn E. Characteristics of telehealth users in NYC for COVID-related care during the coronavirus pandemic. J Am Med Inform Assoc 2020 Dec 09;27(12):1949-1954 [FREE Full text] [doi: 10.1093/jamia/ocaa216] [Medline: 32866249]
33. Jewett PI, Vogel RI, Ghebre R, Hui JY, Parsons HM, Rao A, et al. Telehealth in cancer care during COVID-19: disparities by age, race/ethnicity, and residential status. J Cancer Surviv 2022 Feb;16(1):44-51 [FREE Full text] [doi: 10.1007/s11764-021-01133-4] [Medline: 34800257]
34. Costa F, Janela D, Molinos M, Lains J, Francisco GE, Bento V, et al. Telerehabilitation of acute musculoskeletal multi-disorders: prospective, single-arm, interventional study. BMC Musculoskelet Disord 2022 Jan 04;23(1):29 [FREE Full text] [doi: 10.1186/s12891-021-04891-5] [Medline: 34983488]

35. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Francisco GE, et al. Asynchronous and tailored digital rehabilitation of chronic shoulder pain: a prospective longitudinal cohort study. J Pain Res 2022 Jan 8;15:53-66 [FREE Full text] [doi: 10.2147/JPR.S343308] [Medline: 35035234]
36. Dias Correia F, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Digital versus conventional rehabilitation after total hip arthroplasty: a single-center, parallel-group pilot study. JMIR Rehabil Assist Technol 2019 Jun 21;6(1):e14523 [FREE Full text] [doi: 10.2196/14523] [Medline: 31228176]
37. Gatchel RJ, Mayer TG, Choi Y, Chou R. Validation of a consensus-based minimal clinically important difference (MCID) threshold using an objective functional external anchor. Spine J 2013 Aug;13(8):889-893. [doi: 10.1016/j.spinee.2013.02.015] [Medline: 23523434]
38. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008 Jan 01;33(1):90-94. [doi: 10.1097/BRS.0b013e31815e3a10] [Medline: 18165753]
39. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
41. Ospina MB, Dennett L, Waye A, Jacobs P, Thompson AH. A systematic review of measurement properties of instruments assessing presenteeism. Am J Manag Care 2015 Feb 01;21(2):e171-e185 [FREE Full text] [Medline: 25880491]
42. Xiao J, Bulut O. Evaluating the performances of missing data handling methods in ability estimation from sparse data. Educ Psychol Meas 2020 Oct;80(5):932-954 [FREE Full text] [doi: 10.1177/0013164420911136] [Medline: 32855565]
43. The Burden of Musculoskeletal Diseases in the United States (BMUS). 4th edition. United States Bone and Joint Initiative. URL: http://www.boneandjointburden.org/ [accessed 2021-05-14]
44. Race. United States Census Bureau. URL: https://data.census.gov/cedsci/table?q=United%20States&g=0100000US&tid=DECENNIALPL2020.P1 [accessed 2022-05-14]
45. Iacobucci D. Structural equations modeling: fit indices, sample size, and advanced topics. J Consum Psychol 2009 Oct 03;20(1):90-98 [FREE Full text] [doi: 10.1016/j.jcps.2009.09.003]
46. Brown TA. Confirmatory Factor Analysis for Applied Research. New York, NY, USA: The Guilford Press; 2006.
47. Skúladóttir H, Sveinsdottir H, Holden JE, Gunnarsdóttir TJ, Halldorsdottir S, Björnsdottir A. Pain, sleep, and health-related quality of life after multidisciplinary intervention for chronic pain. Int J Environ Res Public Health 2021 Sep 28;18(19):10233 [FREE Full text] [doi: 10.3390/ijerph181910233] [Medline: 34639534]
48. Farin E, Gramm L, Schmidt E. The patient-physician relationship in patients with chronic low back pain as a predictor of outcomes after rehabilitation. J Behav Med 2013 Jun;36(3):246-258. [doi: 10.1007/s10865-012-9419-z] [Medline: 22476813]
49. Vincent HK, Adams MC, Vincent KR, Hurley RW. Musculoskeletal pain, fear avoidance behaviors, and functional decline in obesity: potential interventions to manage pain and maintain function. Reg Anesth Pain Med 2013;38(6):481-491. [doi: 10.1097/AAP.0000000000000013] [Medline: 24141874]
50. Al-Eisa E. Indicators of adherence to physiotherapy attendance among Saudi female patients with mechanical low back pain: a clinical audit. BMC Musculoskelet Disord 2010 Jun 17;11:124 [FREE Full text] [doi: 10.1186/1471-2474-11-124] [Medline: 20565719]
51. Valaas HL, Klokkerud M, Hildeskår J, Hagland AS, Kjønli E, Mjøsund K, et al. Follow-up care and adherence to self-management activities in rehabilitation for patients with rheumatic and musculoskeletal diseases: results from a multicentre cohort study. Disabil Rehabil (forthcoming) 2021 Nov 30:1-10. [doi: 10.1080/09638288.2021.2008523] [Medline: 34846264]
52. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
53. Gabel CP, Cuesta-Vargas A, Qian M, Vengust R, Berlemann U, Aghayev E, et al. The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain. Eur Spine J 2017 Aug;26(8):2007-2013. [doi: 10.1007/s00586-017-5179-3] [Medline: 28646454]
54. Bijker L, Sleijser-Koehorst ML, Coppieters MW, Cuijpers P, Scholten-Peeters GG. Preferred self-administered questionnaires to assess depression, anxiety and somatization in people with musculoskeletal pain - a modified Delphi study. J Pain 2020;21(3-4):409-417 [FREE Full text] [doi: 10.1016/j.jpain.2019.08.006] [Medline: 31487562]

Abbreviations

CBT: cognitive behavioral therapy
DCP: digital care program
GAD-7: Generalized Anxiety Disorder 7-item scale
LGCA: latent growth curve analysis
MCID: minimum clinically important difference

MSK: musculoskeletal
OR: odds ratio
PHQ-9: Patient Health Questionnaire 9-item
PT: physical therapist
WPAI: Work Productivity and Activity Impairment

Edited by T Leung; submitted 21.07.22; peer-reviewed by A Furlan, G Papandonatos; comments to author 18.08.22; revised version received 14.09.22; accepted 30.09.22; published 31.10.22.

Please cite as:
Scheer J, Costa F, Molinos M, Areias A, Janela D, Moulder RG, Lains J, Bento V, Yanamadala V, Cohen SP, Correia FD
Racial and Ethnic Differences in Outcomes of a 12-Week Digital Rehabilitation Program for Musculoskeletal Pain: Prospective Longitudinal Cohort Study
J Med Internet Res 2022;24(10):e41306
URL: https://www.jmir.org/2022/10/e41306
doi:10.2196/41306
PMID:36189963

Original Paper

The Effectiveness of Patient Training in Inflammatory Bowel Disease Knowledge via Instagram: Randomized Controlled Trial

Dominik Blunck[1*], MSc; Lena Kastner[1], MSc; Michael Nissen[2], MSc; Jacqueline Winkler[3*], MSc

[1]Department of Health Management, Institute of Management, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nuremberg, Germany

[2]Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3]Bristol-Myers Squibb GmbH & Co KGaA, Munich, Germany

[*]these authors contributed equally

Corresponding Author:
Dominik Blunck, MSc
Department of Health Management
Institute of Management
Friedrich-Alexander-Universität Erlangen-Nürnberg
Lange Gasse 20
Nuremberg, 90403
Germany
Phone: 49 911 5302 96394
Email: dominik.blunck@fau.de

Abstract

Background: Patients' knowledge was found to be a key contributor to the success of therapy. Many efforts have been made to educate patients in their disease. However, research found that many patients still lack knowledge regarding their disease. Integrating patient education into social media platforms can bring materials closer to recipients.

Objective: The aim of this study is to test the effectiveness of patient education via Instagram.

Methods: A randomized controlled trial was conducted to test the effectiveness of patient education via Instagram among patients with inflammatory bowel disease. Participants were recruited online from the open Instagram page of a patient organization. The intervention group was educated via Instagram for 5 weeks by the research team; the control group did not receive any educational intervention. The knowledge about their disease was measured pre- and postintervention using the Inflammatory Bowel Disease Knowledge questionnaire. Data were analyzed by comparing mean knowledge scores and by regression analysis. The trial was purely web based.

Results: In total, 49 participants filled out both questionnaires. The intervention group included 25 participants, and the control group included 24 participants. The preintervention knowledge level of the intervention group was reflected as a score of 18.67 out of 24 points; this improved by 3 points to 21.67 postintervention. The postintervention difference between the control and intervention groups was 3.59 points and was statistically significant ($t_{32.88}$=–4.56, 95% CI 1.98-5.19; P<.001). Results of the regression analysis, accounting for preintervention knowledge and group heterogeneity, indicated an increase of 3.33 points that was explained by the intervention (P<.001).

Conclusions: Patient education via Instagram is an effective way to increase disease-related knowledge. Future studies are needed to assess the effects in other conditions and to compare different means of patient education.

Trial Registration: German Clinical Trials Register DRKS00022935; https://tinyurl.com/bed4bzvh

(J Med Internet Res 2022;24(10):e36767) doi:10.2196/36767

KEYWORDS

social media; Instagram; patient training; patient education; disease-related knowledge; RCT; randomized controlled trial; Germany; inflammatory bowel disease; IBD-KNOW

Introduction

Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) is a group of chronic inflammatory diseases of the gastrointestinal tract. IBD can be divided into Crohn disease, ulcerative colitis, and other diseases that present with different gastrointestinal symptoms, such as diarrhea [1]. The global prevalence of IBD is approximately 3.9 million females and 3.0 million males, with a worldwide accelerating incidence [2,3]. The economic burden of IBD is highly relevant. Annual costs per patient were shown to be 3-fold in IBD patients compared to patients without IBD [4]. A systematic review estimated the mean annual health care cost of IBD patients in North America to be over US $13,000 [5]. Although the disease is not yet fully understood [6], there exist different pharmaceutical and nonpharmaceutical interventions. For pharmaceutical interventions, aminosalicylates, corticosteroids, antibiotics, immunomodulative treatments, and different biologic treatments are used, depending on the clinical stage of IBD [7-10]. Nonpharmaceutical interventions are surgery—for example, for patients who are refractory to treatment—and other interventions, such as diets [7]. Because of a greater likelihood of depression or anxiety, resulting in lower quality of life, psychotherapy is a common therapeutic approach as well [7,11-13].

Studies show that IBD patients benefit from higher disease-related knowledge, which has positive effects on the clinical outcomes of their overall therapy [14,15]. Not only in IBD, but also in other, especially chronic, conditions, higher levels of knowledge of the respective condition are related to better outcomes [16,17]. Besides the clinical importance, improving patients' disease-related knowledge is also economically important. A study by Colombara et al [18] found that an increase of 5 points in patients' disease-related knowledge on a 24-point scale could decrease costs in the first year after diagnosis by over €1000.

Disease-Related Knowledge

In this section we describe (1) why higher disease-related knowledge might positively affect clinical outcomes, (2) how other studies approached increasing disease-related knowledge in IBD, (3) how we propose to integrate patient education into patients' daily lives via social media, and (4) how others did so for other indications.

Higher disease-related knowledge has a positive effect on clinical outcomes because it improves adherence and enables shared decision-making, which ultimately leads to better clinical outcomes. Adherence to the treatment plan is a major success factor in therapy. However, in chronic diseases in particular, studies found that medication adherence often is insufficient [19,20]. Higher levels of patient knowledge showed improved adherence in different conditions, for example, because of higher motivation or dispelled misbeliefs [21,22]. Several studies found an improvement in adherence among patients with IBD through different educational interventions and, subsequently, higher rates of knowledge of IBD [23,24]. Bucci et al [25] investigated the factors that predict adherence among Italian patients with IBD and described the complex treatment plan for IBD, which requires taking different pharmaceuticals as well as lifestyle and nutrition changes. Hence, the literature implies a need to enhance knowledge of IBD and related therapies for better adherence. In one study by Elkjaer et al [26], patients with IBD who participated in dedicated educational programs showed better compliance and adherence, higher disease-related knowledge, better quality of life, and better coping with relapsing, leading to a mean relapse duration of 18 days compared to 77 days in the control group. Shared decision-making improves clinical outcomes because therapy plans are aligned with patients' values, lifestyles, and expectations [27-29]. In IBD, shared decision-making is a relevant factor regarding medication therapy [30]. For shared decision-making, however, equitable collaboration between patients and physicians is required. Therefore, high levels of disease-related knowledge are necessary to enable a common understanding of the underlying problems and therapy options [29,31]. Additionally, the majority of patients with IBD also want to be actively involved in the decision-making process, as surveys have shown [32-34], which might be due to high levels of uncertainty associated with IBD [35]. Thus, one important antecedent of shared decision-making is informing patients.

In the case of IBD, different methods to increase disease-related knowledge have been studied. One study compared a telemedicine intervention (ie, SMS text messaging) with standard care (ie, educational materials at clinical appointments) to increase disease-related knowledge in IBD. On a 24-point scale, telemedicine increased the baseline value of 12.6 by 2.4 points, whereas standard care only yielded 1.8 points [36]. In a study where patients received a CD-ROM for self-paced autodidactic learning, participants were able to increase their knowledge from 12.2 points on a 30-point scale to 19.9 points, an increase of 7.6 points. After 9 months of follow-up, the knowledge increase was still 5.3 points higher than at baseline [37]. Another study compared a 12-hour structured education program with standard care (ie, teaching by physician during regular visits). On a 24-point scale, the intervention group's disease-related knowledge increased by 7.71 points immediately after the intervention and 7.94 points after 8 weeks compared to baseline. The control group's disease-related knowledge increased by 3.55 points immediately after standard care and 4.05 points after 8 weeks compared to baseline [38]. In another study, IBD patients were educated through counseling, pill cards, and educational material. In that study, knowledge increased from 8.15 points to 11.65 points [23].

Although different approaches for informing patients have already been studied, they might lack sustainable integration into patients' daily lives. For example, Yin et al [39] argued that most of the educational apps they identified in a scoping review did not proactively inform patients, and patients instead had to access the app by themselves manually; this could be why they were poorly embedded into patients' daily routines. In contrast, social media is discussed as a way to potentially overcome this problem, as many patients already use it and it comes with high interactivity [40].

Therefore, we suggest distributing information via Instagram. Instagram is a widely used social media platform with 1 billion

users worldwide [41]. In most cases, Instagram is accessed via its corresponding smartphone app, which is used to view and share pictures or videos. Users can view pictures and videos in two ways: either via their timeline or the so-called *story* function. Media in the timeline is presented once to the user by the Instagram algorithm but is constantly available. Furthermore, the algorithm orders content as a result of user-based analyses. The story function is found in the top section of the Instagram home screen. Content creators can share short video clips or pictures in the story function, which are then presented to the creator's followers. The order of the stories presented to a user also depends on user-based analyses. Instagram stories are available for 24 hours; however, creators can save their stories using the so-called "Story Highlights" feature, which makes stories constantly available. Buttons to view different categories of highlights are available on every user profile. Besides the sole presentation of pictures or videos in the story, creators can also integrate different interactive functionalities, such as quizzes. A recent study evaluated the use of social media platforms and showed that 59% of Instagram users visited Instagram at least daily, and more than one-third of the users visited the app several times a day [42]. Therefore, it seems like a reasonable approach for integrating patient education into everyday life.

Previous studies of social media–based interventions showed overall good results in improving clinical outcomes and patients' disease-related knowledge about different conditions, for example, diabetes [43]. A review article by Grajales et al [44] reported various approaches for applying social media to health care and patient education. For example, several apps in Facebook are described as well as weblogs. Another paper studied the effect of participation in social health networks on patient activation. Patients with a chronic condition participated in a dedicated social network where they could find medical advice from experts as well as the opportunity to connect with other patients. Higher frequency and duration of usage of this network was associated with higher patient activation, and patients felt more empowered [45].

Aim

The purpose of this study is, thus, to explore whether patient education via Instagram stories is an effective method for educating and informing adult patients with IBD, as compared to patients receiving no intervention, by conducting a randomized controlled trial (RCT).

Methods

Design

This study was conducted as a 2-arm, parallel-group, purely web-based RCT, following the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guideline [46]. The intervention group received disease-related education for 5 weeks, and the control group did not receive any educational treatment. Outcomes were assessed before and after the intervention.

Recruitment and Randomization

For recruitment, we were supported by CHRONISCH GLÜCKLICH e.V., a German patient organization for IBD. The organization owns and operates an Instagram page that had 2332 followers (87.5% female) at the start of recruitment. Comparable pages have similar demographics. They announced the study in their publicly available "Instagram Stories" and called for participation. Participants were included if they met the following inclusion criteria: (1) were older than 18 years of age, (2) had an Instagram account, and (3) were able to fill out a questionnaire. After a recruitment period of 2 weeks, we assigned the participants to either the intervention group or the control group with the help of the online program Research Randomizer [47].

Dropout Effects

According to the intention-to-treat concept, we included all data from all patients in our analysis, whether or not they followed the study protocol [48]. To ensure robustness of our results, we conducted all analyses without dropouts. To better understand dropout effects, we investigated group differences between included participants and those who dropped out with respect to current age, age at diagnosis, sex, diagnosis, and prestudy disease-related knowledge.

Intervention

The intervention group received access to a nonpublic Instagram account, which posted educational material to the story function one to three times per week from June 29, 2020, to July 31, 2020. Furthermore, the stories were saved using the highlights function to be watched later. The posted educational material was either informational or interactive (Figure 1). All educational content was publicly available information about IBD and was reviewed by a physician before being posted by the research team. For interactive purposes, quizzes, for example, were included in the educational stories. Furthermore, participants were not forced or controlled to watch the Instagram stories; they solely received access and followed the account. If participants provided feedback or made requests during the study, such as comments on a story, this was incorporated into successive stories over the 5-week period (ie, higher contrast).

Figure 1. Example screenshots of educational material [content in German].

Outcome Measure

The study's primary outcome was patients' knowledge about IBD. The outcome was measured at baseline (ie, preintervention) and 1 week after the last story was published (ie, postintervention). We measured patients' knowledge by self-assessment using an online questionnaire. There exist different validated questionnaires to measure patients' knowledge about IBD, such as the Crohn's and Colitis Knowledge score [49] and the Inflammatory Bowel Disease Knowledge (IBD-KNOW) questionnaire [50]. We chose the IBD-KNOW questionnaire because it is newer and includes a broader field of disease and therapy-related knowledge, such as biologics. We measured the patients' knowledge about IBD by using the validated IBD-KNOW questionnaire. For this purpose, we translated the original English-language questionnaire into German (Multimedia Appendix 1). This translated version was reviewed by a physician. The questionnaire consists of 24 items, asking questions about IBD facts with response options of "true," "false," and "I don't know." The number of correct answers—"I don't know" is not counted as correct—represents the respondent's level of knowledge about IBD and, hence, the score ranges from 0 to 24 points. The online questionnaire was evaluated by application of the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [51].

Besides the 24 IBD-specific questions, we included several sociodemographic and disease-related variables in the questionnaire, which were included as control variables in the regression analyses.

Sample Size

To identify the required sample size, we performed a power analysis. An improvement of 3 points in the IBD-KNOW score has been previously regarded as clinically important [36,52]. At an SD of 4.7 [15] and to detect group differences of at least 3 points on the IBD-KNOW scale, with power greater than 0.8 and $\alpha<.05$, a sample size of 40 participants per group was required [53]. We anticipated a dropout rate of 20%, giving a total planned sample size of 100 participants.

Statistical Analysis

Overview

We analyzed the study's data in three ways. Firstly, we descriptively analyzed the study participants' characteristics. Secondly, we conducted inferential statistics to display group and time differences in level of knowledge. Thirdly, we conducted a regression analysis. All statistical analyses were performed with R statistical software (version 4.0.0; R Foundation for Statistical Computing) [54,55]. We used the following R packages: pwr for power calculation [53], ggplot2 for data visualization [56], car for calculating variance inflation factors [57], and dplyr and tidyr for data management [58,59]. *P* values of less than .05 were considered statistically significant.

Inferential Statistics

To analyze group differences regarding categorical variables, we used the chi-square test. For continuous variables, we conducted the Welch *t* test.

Regression Analysis

To further analyze the effects, account for group heterogeneity, and ensure robustness of our results, we estimated an ordinary least squares (OLS) regression model of patients' knowledge with a difference-in-differences approach (ie, lm()-function in R).

The dependent variable in the regression model was the IBD-KNOW score. The independent variables included a group dummy variable, a time dummy variable, and an interaction term of group and time. The group dummy value was 1 for the treatment group and 0 for the control group; the time dummy value was 1 for the postintervention questionnaire and 0 for the preintervention questionnaire. The covariates were chosen to control for further effects that are associated with learning. Hence, we controlled for sex (dummy variable, female = 1), age in years, the duration in years that the patient has lived with their IBD diagnosis at the time of the study (ie, current age – age at diagnosis), and diagnosis (dummy variable for Crohn disease) [60,61]. This is reflected in the following equation:

$$y = \beta_0 + \beta_1 dSex + \beta_2 dDiagnosis + \beta_3 Age + \beta_4 Duration + \beta_5 dTime + \beta_6 dGroup + \beta_7 (dTime \times dGroup) + e$$

In the regression analysis, we followed the intention-to-treat approach by including all dropouts in the analysis. However, we estimated further models with dropouts excluded to ensure robustness of the results. Multicollinearity was checked by calculating variance inflation factors. Values greater than 5 were considered to indicate multicollinearity [62].

Ethics Approval

This study was prospectively approved by the Ethics Committee of Friedrich-Alexander-Universität Erlangen-Nürnberg (reference No. 202_20 B) and retrospectively registered in the German Clinical Trials Register (DRKS00022935). All participants declared informed consent before the study after receiving patient information and the data privacy declaration.

Results

Out of 83 initial participants, 40 (48%) were assigned to the control group and 43 (52%) were assigned to the treatment group. In total, 15 participants from the control group and 19 from the intervention group were lost to follow-up because they did not fill out both questionnaires and were, thus, regarded as dropouts. This left a total of 49 participants—25 (51%) in the control group and 24 (49%) in the intervention group—who were analyzed (Figure 2). However, all outcome analyses are reported with and without dropouts in this section. The characteristics of the intervention and control group participants are displayed in Table 1; we did not find statistically significant differences between the control and intervention groups.

We did not find significant group differences between the included participants and the dropout group with respect to age at diagnosis (P=.34), sex (P=.37), type of diagnosis (P=.93), and prestudy IBD knowledge (P=.17). A difference in age between the dropout group and the included participants was found (P=.04), with the dropouts being 3 years older on average. This difference in age did not yield a difference regarding the length of IBD history, which is the difference between current age and age at diagnosis (P=.27).

Without excluding dropouts (ie, intention-to-treat approach), preintervention knowledge in the control group was reflected by a mean of 17.73 (SD 3.72) points, and preintervention knowledge in the intervention group was reflected by a mean of 18.33 (SD 3.13) points; the difference was not statistically significant ($t_{76.47}$=–0.79, 95% CI –2.11 to 0.91; P=.43). When dropouts were excluded, preintervention knowledge in the whole sample was reflected by a mean of 18.47 (SD 3.40) points. With dropouts excluded, preintervention knowledge in the control group was reflected by a mean of 18.28 (SD 3.76) points, and preintervention knowledge in the intervention group was reflected by a mean of 18.67 (SD 3.05) points. The difference between the control and intervention groups before the intervention was not statistically significant ($t_{45.73}$=–0.40, 95% CI –2.35 to 1.58; P=.69). Postintervention knowledge was reflected by a mean of 18.08 (SD 3.60) points in the control group and 21.67 (SD 1.55) points in the intervention group. This difference of 3.59 points was statistically significant ($t_{32.88}$=–4.56, 95% CI –5.19 to –1.98; power=0.99; P<.001). The pre- and postintervention knowledge levels by the control and intervention groups are displayed in Figure 3.

Figure 2. Flowchart of participants.

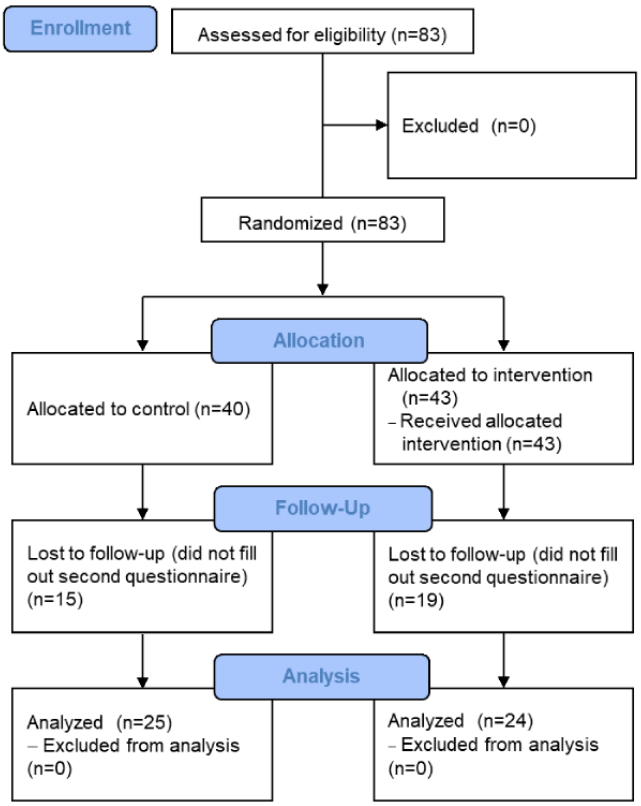


Table 1. Characteristics of the study participants.

Characteristics	Control group (n=25)	Intervention group (n=24)	Full sample (N=49)	*t* test[a] (*df*)	χ^{2a} (*df*)	*P* value
Age (years), mean (SD)	25.88 (5.82)	26.96 (6.69)	26.41 (6.22)	–0.60 (45.53)	N/A[b]	.55
Age at diagnosis (years), mean (SD)	21.40 (6.42)	19.88 (8.07)	20.65 (7.24)	0.73 (43.91)	N/A	.47
Sex, n (%)						
Female	23 (92)	24 (100)	47 (96)	N/A	0.5 (1)	.49
Male	2 (8)	0 (0)	2 (4)	N/A	__[c]	—
Type of diagnosis, n (%)						
Crohn disease	14 (56)	16 (67)	30 (61)	N/A	0.2 (1)	.64
Ulcerative colitis	11 (44)	8 (33)	19 (39)	N/A	—	—
Knowledge about IBD[d], IBD-KNOW[e] score, mean (SD)						
Preintervention	18.28 (3.76)	18.67 (3.05)	18.47 (3.40)	–0.40 (45.73)	N/A	.69
Postintervention	18.08 (3.60)	21.67 (1.55)	19.84 (3.31)	–4.56 (32.88)	N/A	<.001

[a]The *t* test (2-tailed) and chi-square test were used to measure the difference between the control and intervention groups.

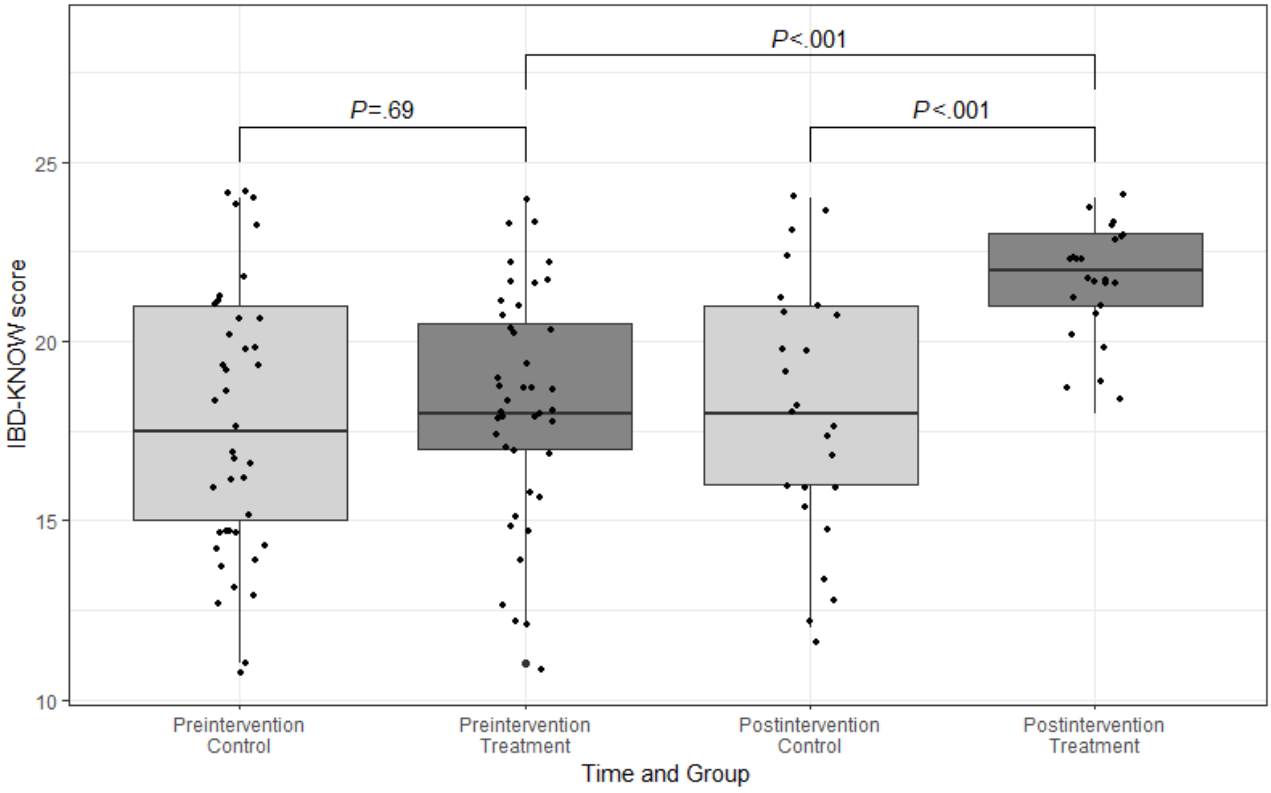
[b]N/A: not applicable; this test was not applied to this variable.

[c]The chi-square value and its related *P* value for a group are reported in the top row for that group.

[d]IBD: inflammatory bowel disease.

[e]IBD-KNOW: Inflammatory Bowel Disease Knowledge; scores range from 0 to 24 points.

Figure 3. Levels of pre- and postintervention knowledge by control and intervention groups. IBD-KNOW: Inflammatory Bowel Disease Knowledge.



The results of the OLS regression analysis are displayed in Table 2. Model 1 shows the baseline effect of the selected control variables on patients' knowledge scores. Model 2 adds the time and group dummy variables, as well as the interaction term for these variables. The variable of interest is the interaction term, as it describes the main treatment effect. Patients in the treatment group increased their knowledge score by 3.07 points, all other things being equal, compared to the control group (*P*=.001; see Multimedia Appendix 2 for a visualization of the treatment effect).

R^2 represents the proportion of variance in the dependent variable that is explained by the model. In model 1, 1% of the

variance in patients' knowledge is explained by the control variables. After adding the independent variables, the R^2 of model 2 shows that 18% of the variance of patients' knowledge is explained by the variables. The adjusted R^2, which considers the number of control variables, in model 2 indicates that 13% of the variance is explained by model 2, a gain of 15 percentage points (pp) compared to model 1. The statistically significant *F* test values in model 2 indicate an overall significant model [61].

Variance inflation factors were all well below the cutoff of 5, with a maximum in model 1 of 1.06 in age and a maximum in model 2 of 2.57 in the interaction term; this was expected, as the interaction was a linear combination of two other variables. Given these results, we do not consider multicollinearity to be a major problem in our analysis.

Results of the robustness test, including the participants who dropped out, confirmed our results: estimate of time × treatment = 3.21 (*P*=.01); adjusted R^2=0.17; $F_{7,90}$=3.83 (*P*=.001). The results can be found in Table S1 in Multimedia Appendix 3.

Qualitative feedback from participants was incorporated during the study. For example, participants noted that some story slides were difficult to read, as IBD can affect patients' eyes. Therefore, story slides were designed in high contrast after this feedback. Furthermore, we received a lot of positive feedback. Participants regarded the interventions as useful and meaningful. They also noted that they learned a lot—especially newly diagnosed participants—and stated that these interventions should be much more common.

Table 2. Difference-in-differences regression of knowledge about inflammatory bowel disease.

Variables and measures	Model 1		Model 2	
	Value	*P* value	Value	*P* value
Control variables, estimated β coefficient (SE)				
Constant	19.82 (1.87)	<.001	19.72 (1.77)	<.001
Female	–0.68 (1.29)	.60	–1.45 (1.19)	.23
Crohn disease[a]	0.34 (0.63)	.59	0.16 (0.59)	.79
Ulcerative colitis[a]	Reference	—[b]	Reference	—
Age	–0.03 (0.05)	.49	–0.03 (0.05)	.53
Duration	0.04 (0.06)	.51	0.00 (0.05)	.95
Independent variables, estimated β coefficient (SE)				
Time[a]	N/A[c]	N/A	0.33 (0.83)	.69
Intervention[a]	N/A	N/A	0.64 (0.72)	.38
Time × intervention	N/A	N/A	3.07 (1.17)	.001
Observations, n	132	—	132	—
R^2	0.01	—	0.18	—
Delta R^2	N/A	N/A	0.17	—
Adjusted R^2	–0.02	—	0.13	—
Delta adjusted R^2	N/A	N/A	0.15	—
F test (*df*)	0.320 (4, 127)	.86	3.889 (7, 124)	<.001

[a]Dummy variable.

[b]Not calculated.

[c]N/A: not applicable; model 1 was not applied to these variables or measures.

Discussion

Principal Findings

To answer the research question of whether educating adult patients with IBD via Instagram is effective, we conducted an RCT in a sample of 49 participants. After 5 weeks of training via Instagram stories, the intervention group yielded statistically significant and relatively higher levels of disease-related knowledge. Therefore, this study provides evidence for the effectiveness of patient education via Instagram.

With a mean of 76.95% correct answers (mean score of 18.47 out of 24), our sample showed an already-high mean knowledge level at baseline, compared to other studies in this area. For example, Abutaleb et al [36] found 52.50% correct answers during the preintervention stage. Others found mean baseline knowledge levels of 26.67% (8/30) [63], 33.33% (8/24) [18], 40.67% (12.2/30) [37], 40.79% (9.79/24) and 48.25% (11.58/24)

[38], and 62.90% (18.87/30) [64]. Along with the relatively high baseline knowledge level, our study showed an increase in mean disease-related knowledge by 12.50 pp. Other studies achieved increases of 10 pp with telemedicine and 7.5 pp with standard interventions [36], 25.33 pp with a CD-ROM program [37], and 32.13 pp with a formal education program and 14.79 pp with a standard intervention [38]. Hence, the knowledge increase presented in our study is on the lower bound compared to other interventions. However, the study designs are not comparable without restrictions, for example, because of different intensity and frequency of interventions. Furthermore, higher baseline values come with less improvement from educational interventions [36], which is reasonable due to a saturation effect and a natural upper limit of the knowledge scale.

The dropout rate in this study was 41% (34/83) and was, thus, relatively high compared to other studies; for example, one study found 25% loss to follow-up after 6 months and 26% loss to follow-up after 12 months [36], whereas another study found 16% dropout immediately after the intervention and 22% loss to follow-up after 8 weeks [38]. We believe that the high dropout rate in our study may be due to the fact that, in order to prevent forced results, we did not send reminders to the participants to complete the questionnaires. Although the dropout group did not differ from the included participants regarding parameters such as length of IBD history or prestudy knowledge, dropouts were significantly older than included participants. A reason for this observation might be that older patients might have lower computer literacy and, thus, were more likely to drop out. Hence, future studies could address this issue in further elaborating the interplay of age and learning via social media in patients with IBD.

The unexpectedly high dropout rate ultimately led to a relatively low number of participants. This was not in line with the assumptions used for the power analysis. Future studies should take measures to either (1) expect a higher dropout rate and recruit a larger number of participants or (2) decrease the overall dropout rate. The latter may be achieved by using reminders or incentives. We did not take these measures in our study in order to reduce bias.

Finally, we found a high proportion of women among the followers of the organization specific to patients with IBD on Instagram. This may suggest that men generally have different coping strategies for dealing with IBD than women.

Contribution

To our knowledge, this was the first study to analyze the effect of patient training via Instagram on patients' disease-related knowledge. One main contribution of our study is evidence for the effectiveness of patient education via Instagram. Future work in this area should focus on disseminating educational content in regular care. One major challenge for this could be quality assurance because everybody could publish apparent educational content without expert review. If health care providers actively use social media platforms in the future, a high level of quality in educational material could be ensured. Another challenge might be the long-term motivation of users. Potential ways to reduce retention issues are high-quality content, high levels of monitoring and interaction, or the use of Instagram ads to increase visibility. However, the latter mechanism, in particular, might bias results in the study setting and would be more suitable in a regular care setting.

A difference in this study compared to previous studies is that participants in this study did not participate in dedicated trainings. This means that patients only received access to the Instagram account and were responsible for watching or actively participating. In classical patient educational interventions [38], patients actively participate in a training session, a physician visit, or similar. As it is not feasible in a regular care setting to ensure continued training via dedicated trainings, we contributed by providing a solution that is integrated into patients' daily routines, without a cost to health care providers, and that can be used on a long-term and continued basis. Once educational material is designed and conceptualized, it could be used and reused in a large patient population. Compared to other, previously mentioned, ways of increasing patients' disease-related knowledge, our approach is easy to implement, comes with good scalability, integrates educational content into patients' lives, and addresses young people in particular. Furthermore, the proposed approach allows possibilities for patient organizations to closer engage with patients. Another application of educational social media interventions is the education of patients' friends and family members. As those people are often affected or involved in the care of patients with chronic conditions, higher disease-related knowledge among friends and family members could also increase their understanding of patients' situations and therapies, which subsequently would support patients. Furthermore, we contributed by providing a German translation of the IBD-KNOW questionnaire.

Limitations

Our study comes with several limitations. First, patient recruitment took place via the Instagram pages of a German patient organization. This might bias and underestimate results for the total relevant population because we assumed that the patient organization's Instagram page was being followed by an already-interested audience. For example, studies found that patients who are members of a patient organizations yield higher knowledge scores than patients who are not [50]. Therefore, the knowledge levels of this respective sample might already be above average. On the other hand, however, one could argue that the sample of patients could be more highly motivated and have a higher willingness to learn due to their higher level of interest, which counteracts this effect. Additionally, the study setting may have led to another selection bias because young and computer-literate people, in particular, are Instagram users, which limits generalizability. Another limitation might arise from dropouts. As 34 participants were lost to follow-up, our overall findings might be biased if the dropout probability was associated with the knowledge score, specifically with learning. Due to the unexpectedly high dropout rate, the sample size of our study was relatively small. Inclusion of larger study populations might be beneficial in gaining a better understanding of our findings.

Additionally, participants in this study were almost exclusively female. As the proportion of women among all patients with IBD is much lower [2], the generalizability of this study to the whole IBD population may be limited. However, the high proportion of women in our study is due to the demographic composition of Instagram followers of the patient organization with which we collaborated for recruitment.

To assert the sustainability of the effect of education via Instagram, further studies with a longer follow-up period are needed. However, the real-life setting of the proposed educational mode has a continuous character. This means that patients have continuous access to the educational material instead, for example, of a one-time visit at a seminar, which rather reduces the need for follow-up studies. Furthermore, previous studies found that the knowledge increase gained by patients with IBD stays relatively constant over time [37,38].

Additionally, we only considered German patients, which might reduce the generalizability of our results. Studies show that knowledge levels differ between countries [65]. Future studies should, therefore, focus on multicenter study designs or evaluate results across countries.

The interest in the educational material in our study might be higher than in a real-life setting because of a trial effect. Patients might be interested more or might learn more because they know they are part of a study [66] and not blinded. Therefore, the effect might be overestimated. To validate the effectiveness of patient education via Instagram or other social media channels, further research (eg, observational studies) is needed.

Future Research

This study recommends different questions for future research. First, patient education via Instagram or other social media should be directly compared with other means of patient education, in order to compare effectiveness in a head-to-head comparison. Second, the effectiveness of Instagram patient education should be tested in other chronic conditions as well. Third, the economic effects of patient education via Instagram—or social media in general—should be explored. Integration into patients' daily routines might reduce costs for transportation to a training facility or physician. Additionally, patient education via social media, such as Instagram, is easy to scale and increases accessibility, which leads to lower costs at training facilities or for physicians. Fourth, before rolling out Instagram patient education in regular settings, quality requirements should be defined to enable systematic dissemination and prevent communication of misleading or false information to patients.

Conclusions

To test the effectiveness of patient training via Instagram, we conducted an RCT with 49 patients with IBD. The intervention group received access to an Instagram account, which posted educational material over 5 weeks. The outcome—patients' knowledge about IBD—was measured at the pre- and postintervention stages using a questionnaire whose response scores ranged from 0 to 24 points. The intervention group yielded 3.59 more points than the control group, on average, after the intervention (P<.001), with no significant differences before the intervention. Therefore, we conclude that Instagram is an effective tool for educating patients and demonstrates large potential for future support of chronic conditions.

Acknowledgments

We would like to thank CHRONISCH GLÜCKLICH e.V. for supporting the recruitment of participants. We acknowledge financial support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg within the funding program Open Access Publication Funding.

Authors' Contributions

JW was responsible for conceptualization of the study, methodology, investigation, and writing, reviewing, and editing the paper. DB was responsible for conceptualization of the study; methodology; formal analysis; writing the original draft; reviewing and editing subsequent drafts; visualization; supervision; and project administration. LK and MN were responsible for writing the original draft and for reviewing and editing subsequent drafts.

Conflicts of Interest

JW worked during her master's course—where this paper was initiated as a student project—for Roche Pharma AG as a patient partnership manager in the field of gastroimmunology. Roche did not provide any funding for the study and had no influence in initiating, planning, designing, and conducting the study; collection, analysis, or interpretation of data; the writing of the manuscript; or the decision to publish the paper.

Multimedia Appendix 1
Inflammatory Bowel Disease Knowledge questionnaire.
[DOCX File , 18 KB - jmir_v24i10e36767_app1.docx]

Multimedia Appendix 2
Visualization of the treatment effect.

[PNG File , 6 KB - jmir_v24i10e36767_app2.png]

Multimedia Appendix 3
Difference-in-differences regression results (robustness check).
[DOCX File , 17 KB - jmir_v24i10e36767_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1184 KB - jmir_v24i10e36767_app4.pdf]

References

1. Guan Q. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. J Immunol Res 2019;2019:7247238 [FREE Full text] [doi: 10.1155/2019/7247238] [Medline: 31886308]
2. GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2020 Jan;5(1):17-30 [FREE Full text] [doi: 10.1016/S2468-1253(19)30333-4] [Medline: 31648971]
3. Liu S, Zhao W, Lan P, Mou X. The microbiome in inflammatory bowel diseases: From pathogenesis to therapy. Protein Cell 2021 May;12(5):331-345 [FREE Full text] [doi: 10.1007/s13238-020-00745-3] [Medline: 32601832]
4. Park KT, Ehrlich OG, Allen JI, Meadows P, Szigethy EM, Henrichsen K, et al. Inflamm Bowel Dis 2020 Jan 01;26(1):1-10 [FREE Full text] [doi: 10.1093/ibd/izz104] [Medline: 31112238]
5. van Linschoten RCA, Visser E, Niehot CD, van der Woude CJ, Hazelzet JA, van Noord D, et al. Systematic review: Societal cost of illness of inflammatory bowel disease is increasing due to biologics and varies between continents. Aliment Pharmacol Ther 2021 Aug;54(3):234-248 [FREE Full text] [doi: 10.1111/apt.16445] [Medline: 34114667]
6. Shah SC, Khalili H, Gower-Rousseau C, Olen O, Benchimol EI, Lynge E, et al. Sex-based differences in incidence of inflammatory bowel diseases-pooled analysis of population-based studies from western countries. Gastroenterology 2018 Oct;155(4):1079-1089.e3. [doi: 10.1053/j.gastro.2018.06.043] [Medline: 29958857]
7. Lamb CA, Kennedy NA, Raine T, Hendy PA, Smith PJ, Limdi JK, IBD guidelines eDelphi consensus group, et al. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. Gut 2019 Dec;68(Suppl 3):s1-s106 [FREE Full text] [doi: 10.1136/gutjnl-2019-318484] [Medline: 31562236]
8. Asgharpour A, Cheng J, Bickston SJ. Adalimumab treatment in Crohn's disease: An overview of long-term efficacy and safety in light of the EXTEND trial. Clin Exp Gastroenterol 2013 Aug 30;6:153-160 [FREE Full text] [doi: 10.2147/CEG.S35163] [Medline: 24039442]
9. Kruis W, Kiudelis G, Rácz I, Gorelov IA, Pokrotnieks J, Horynski M, International Salofalk OD Study Group. Once daily versus three times daily mesalazine granules in active ulcerative colitis: A double-blind, double-dummy, randomised, non-inferiority trial. Gut 2009 Feb;58(2):233-240 [FREE Full text] [doi: 10.1136/gut.2008.154302] [Medline: 18832520]
10. Lim W, Wang Y, MacDonald JK, Hanauer S. Aminosalicylates for induction of remission or response in Crohn's disease. Cochrane Database Syst Rev 2016 Jul 03;7:CD008870 [FREE Full text] [doi: 10.1002/14651858.CD008870.pub2] [Medline: 27372735]
11. Mikocka-Walus A, Knowles SR, Keefer L, Graff L. Controversies revisited: A systematic review of the comorbidity of depression and anxiety with inflammatory bowel diseases. Inflamm Bowel Dis 2016 Mar;22(3):752-762. [doi: 10.1097/MIB.0000000000000620] [Medline: 26841224]
12. Knowles SR, Graff LA, Wilding H, Hewitt C, Keefer L, Mikocka-Walus A. Quality of life in inflammatory bowel disease: A systematic review and meta-analyses-Part I. Inflamm Bowel Dis 2018 Mar 19;24(4):742-751. [doi: 10.1093/ibd/izx100] [Medline: 29562277]
13. Knowles SR, Keefer L, Wilding H, Hewitt C, Graff LA, Mikocka-Walus A. Quality of life in inflammatory bowel disease: A systematic review and meta-analyses-Part II. Inflamm Bowel Dis 2018 Apr 23;24(5):966-976. [doi: 10.1093/ibd/izy015] [Medline: 29688466]
14. Norton BA, Thomas R, Lomax KG, Dudley-Brown S. Patient perspectives on the impact of Crohn's disease: Results from group interviews. Patient Prefer Adherence 2012;6:509-520 [FREE Full text] [doi: 10.2147/PPA.S32690] [Medline: 22879737]
15. Park J, Yoon H, Shin CM, Park YS, Kim N, Lee DH. Higher levels of disease-related knowledge reduce medical acceleration in patients with inflammatory bowel disease. PLoS One 2020;15(6):e0233654 [FREE Full text] [doi: 10.1371/journal.pone.0233654] [Medline: 32502199]
16. Feng YY, Chaves GS, Shi W, Pakosh M, Zhang L, Gallagher R, et al. Education interventions in Chinese cardiac patients on health behaviours, disease-related knowledge, and health outcomes: A systematic review and meta-analysis. Patient Educ Couns 2021 May;104(5):1018-1029. [doi: 10.1016/j.pec.2020.12.001] [Medline: 33349505]

17. Marciano L, Camerini A, Schulz PJ. The role of health literacy in diabetes knowledge, self-care, and glycemic control: A meta-analysis. J Gen Intern Med 2019 Jun;34(6):1007-1017 [FREE Full text] [doi: 10.1007/s11606-019-04832-y] [Medline: 30877457]
18. Colombara F, Martinato M, Girardin G, Gregori D. Higher levels of knowledge reduce health care costs in patients with inflammatory bowel disease. Inflamm Bowel Dis 2015 Mar;21(3):615-622. [doi: 10.1097/MIB.0000000000000304] [Medline: 25636120]
19. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
20. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: Meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-887.e1. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]
21. Bain-Brickley D, Butler LM, Kennedy GE, Rutherford GW. Interventions to improve adherence to antiretroviral therapy in children with HIV infection. Cochrane Database Syst Rev 2011 Dec 07(12):CD009513 [FREE Full text] [doi: 10.1002/14651858.CD009513] [Medline: 22161452]
22. George M. Adherence in asthma and COPD: New strategies for an old problem. Respir Care 2018 Jun;63(6):818-831 [FREE Full text] [doi: 10.4187/respcare.05905] [Medline: 29794213]
23. Ashok K, Mathew AA, Thomas A, Mohan D, Gopalakrishna R, Reghu R. Clinical pharmacist's interventions on medication adherence and knowledge of inflammatory bowel disease patients. J Young Pharm 2017 Jul 01;9(3):381-385. [doi: 10.5530/jyp.2017.9.76]
24. Lim JK, Lee YJ, Park JH. Medication-related knowledge and medication adherence in pediatric and adolescent patients with inflammatory bowel disease. J Korean Med Sci 2020 Apr 13;35(14):e92 [FREE Full text] [doi: 10.3346/jkms.2020.35.e92] [Medline: 32281312]
25. Bucci C, Zingone F, Tammaro S, Iovino P, Santonicola A, Ciacci C. Factors predicting the adherence to the therapy of Italian IBD patients. Gastroenterol Res Pract 2017;2017:6719345 [FREE Full text] [doi: 10.1155/2017/6719345] [Medline: 28848602]
26. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: A randomised controlled trial of the web-guided 'Constant-care' approach. Gut 2010 Dec;59(12):1652-1661. [doi: 10.1136/gut.2010.220160] [Medline: 21071584]
27. Hargraves I, LeBlanc A, Shah ND, Montori VM. Shared decision making: The need for patient-clinician conversation, not just information. Health Aff (Millwood) 2016 Apr;35(4):627-629. [doi: 10.1377/hlthaff.2015.1354] [Medline: 27044962]
28. Backman WD, Levine SA, Wenger NK, Harold JG. Shared decision-making for older adults with cardiovascular disease. Clin Cardiol 2020 Feb;43(2):196-204 [FREE Full text] [doi: 10.1002/clc.23267] [Medline: 31580493]
29. Fox JC, Lipstein EA. Shared decision making in gastroenterology: Challenges and opportunities. Mayo Clin Proc Innov Qual Outcomes 2020 Apr;4(2):183-189 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2019.11.003] [Medline: 32280929]
30. Siegel CA. Shared decision making in inflammatory bowel disease: Helping patients understand the tradeoffs between treatment options. Gut 2012 Mar;61(3):459-465. [doi: 10.1136/gutjnl-2011-300988] [Medline: 22187072]
31. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: A model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
32. Baars JE, Markus T, Kuipers EJ, van der Woude CJ. Patients' preferences regarding shared decision-making in the treatment of inflammatory bowel disease: Results from a patient-empowerment study. Digestion 2010;81(2):113-119. [doi: 10.1159/000253862] [Medline: 20093836]
33. Conrad S, Hüppe A, Raspe H. [Preference of patients with inflammatory bowel disease regarding information and shared decision-making: Results from a cross-sectional survey in Germany] [Article in German]. Z Gastroenterol 2012 Apr;50(4):364-372. [doi: 10.1055/s-0031-1281949] [Medline: 22467539]
34. Baars JE, Siegel CA, van't Spijker A, Markus T, Kuipers EJ, van der Woude CJ. Inflammatory bowel disease-patients are insufficiently educated about the basic characteristics of their disease and the associated risk of colorectal cancer. Dig Liver Dis 2010 Nov;42(11):777-784. [doi: 10.1016/j.dld.2010.03.023] [Medline: 20472518]
35. Barberio B, Zamani M, Black CJ, Savarino EV, Ford AC. Prevalence of symptoms of anxiety and depression in patients with inflammatory bowel disease: A systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2021 May;6(5):359-370. [doi: 10.1016/S2468-1253(21)00014-5] [Medline: 33721557]
36. Abutaleb A, Buchwald A, Chudy-Onwugaje K, Langenberg P, Regueiro M, Schwartz DA, et al. Inflammatory bowel disease telemedicine clinical trial: Impact of educational text messages on disease-specific knowledge over 1 year. Inflamm Bowel Dis 2018 Sep 15;24(10):2191-2197 [FREE Full text] [doi: 10.1093/ibd/izy149] [Medline: 29788137]
37. Boamah LM, Bohren JR, Pentiuk S, Baker R, Yi M, Moyer MS. Development and testing of a CD-ROM program for improving adolescent knowledge of inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2010 May;50(5):521-525. [doi: 10.1097/MPG.0b013e3181b1356a] [Medline: 20639710]
38. Waters BM, Jensen L, Fedorak RN. Effects of formal education for patients with inflammatory bowel disease: A randomized controlled trial. Can J Gastroenterol 2005 Apr;19(4):235-244. [doi: 10.1155/2005/250504] [Medline: 15861266]

39. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: Scoping review. J Med Internet Res 2019 Aug 19;21(8):e14630 [FREE Full text] [doi: 10.2196/14630] [Medline: 31429410]
40. Markham MJ, Gentile D, Graham DL. Social media for networking, professional development, and patient engagement. Am Soc Clin Oncol Educ Book 2017;37:782-787 [FREE Full text] [doi: 10.1200/EDBK_180077] [Medline: 28561727]
41. About us. Instagram. URL: https://about.instagram.com/about-us [accessed 2021-03-25]
42. 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 2022-10-10]
43. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: Rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
44. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: A review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
45. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent surfing on social health networks is associated with increased knowledge and patient health activation. J Med Internet Res 2016 Aug 10;18(8):e212 [FREE Full text] [doi: 10.2196/jmir.5832] [Medline: 27511272]
46. 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]
47. Urbanik GC, Plous S. Research Randomizer. URL: http://www.randomizer.org/ [accessed 2020-06-06]
48. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
49. Eaden JA, Abrams K, Mayberry JF. The Crohn's and Colitis Knowledge Score: A test for measuring patient knowledge in inflammatory bowel disease. Am J Gastroenterol 1999 Dec;94(12):3560-3566. [doi: 10.1111/j.1572-0241.1999.01536.x] [Medline: 10606319]
50. Yoon H, Yang S, So H, Lee KE, Park SH, Jung S, et al. Development, validation, and application of a novel tool to measure disease-related knowledge in patients with inflammatory bowel disease. Korean J Intern Med 2019 Jan;34(1):81-89 [FREE Full text] [doi: 10.3904/kjim.2017.104] [Medline: 29172400]
51. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
52. Hou JK, Turkeltaub JA, McCarty III TR, El-Serag HB. Assessment of disease specific knowledge and health-related quality of life among United States military veterans with inflammatory bowel disease. World J Gastroenterol 2015 May 21;21(19):6001-6007 [FREE Full text] [doi: 10.3748/wjg.v21.i19.6001] [Medline: 26019466]
53. Champely S, Ekstrom C, Dalgaard P. pwr: Basic functions for power analysis. The Comprehensive R Archive Network. 2020 Mar 17. URL: https://cran.r-project.org/package=pwr [accessed 2022-10-10]
54. R Core Team. R (version 4.0.0). The R Project for Statistical Computing. 2020. URL: https://www.r-project.org [accessed 2022-10-10]
55. RStudio Team. RStudio (version 1.2.5042). RStudio. 2020. URL: https://www.rstudio.com [accessed 2022-10-10]
56. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd edition. Cham, Switzerland: Springer; 2016.
57. Fox J, Weisberg S. An R Companion to Applied Regression. 3rd edition. Thousand Oaks, CA: SAGE Publications, Inc; 2018.
58. Wickham H, Francois R, Henry L, Müller K. dplyr: A grammar of data manipulation. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/package=dplyr [accessed 2022-10-10]
59. Wickham H, Girlich M. tidyr: Tidy messy data. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/package=tidyr [accessed 2022-10-10]
60. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. Contemp Clin Trials Commun 2018 Jun;10:80-85 [FREE Full text] [doi: 10.1016/j.conctc.2018.03.008] [Medline: 29696162]
61. Wooldridge JM. Introductory Econometrics: A Modern Approach. 6th edition. Boston, MA: Cengage; 2016.
62. Sheather SJ. A Modern Approach to Regression With R. New York, NY: Springer; 2009.
63. Butcher RO, Law TL, Prudham RC, Limdi JK. Patient knowledge in inflammatory bowel disease: CCKNOW, how much do they know? Inflamm Bowel Dis 2011 Oct;17(10):E131-E132. [doi: 10.1002/ibd.21810] [Medline: 21739535]
64. Moradkhani A, Kerwin L, Dudley-Brown S, Tabibian JH. Disease-specific knowledge, coping, and adherence in patients with inflammatory bowel disease. Dig Dis Sci 2011 Oct;56(10):2972-2977. [doi: 10.1007/s10620-011-1714-y] [Medline: 21538016]
65. Kim JY, Yoon H, Hwang JS, Yang S, Park SH, Loftus EV. Comparison of disease-related knowledge of patients with inflammatory bowel disease between the West and the East using an updated questionnaire (IBD-KNOW). J Clin Gastroenterol 2020 Sep;54(8):720-724. [doi: 10.1097/MCG.0000000000001283] [Medline: 31764490]
66. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ 2015 Sep 04;351:h4672. [doi: 10.1136/bmj.h4672] [Medline: 26341898]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
IBD: inflammatory bowel disease
IBD-KNOW: Inflammatory Bowel Disease Knowledge
OLS: ordinary least squares
pp: percentage points
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 16.02.22; peer-reviewed by S Esworthy, FH Leung; comments to author 29.07.22; revised version received 12.09.22; accepted 19.09.22; published 19.10.22.

Please cite as:
Blunck D, Kastner L, Nissen M, Winkler J
The Effectiveness of Patient Training in Inflammatory Bowel Disease Knowledge via Instagram: Randomized Controlled Trial
J Med Internet Res 2022;24(10):e36767
URL: https://www.jmir.org/2022/10/e36767
doi:10.2196/36767
PMID:36260385

Original Paper

Exploring Social Support in an Online Support Community for Tourette Syndrome and Tic Disorders: Analysis of Postings

Mercédesz Judit Soós[1], BSc, MSc; Neil S Coulson[1], MA, MSc, PhD; E Bethan Davies[2,3], BSc, MSc, PhD

[1]Population and Lifespan Sciences, School of Medicine, University of Nottingham, Queen's Medical Centre, Nottingham, United Kingdom

[2]National Institute for Health and Care Research MindTech MedTech Co-operative, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[3]Clinical Neurosciences and Mental Health, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
E Bethan Davies, BSc, MSc, PhD
National Institute for Health and Care Research MindTech MedTech Co-operative
Institute of Mental Health
School of Medicine, University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 0015 748 4238
Email: bethan.davies@nottingham.ac.uk

Abstract

Background: Online support communities have become an accessible way of gaining social, emotional, and informational support from peers and may be particularly useful for individuals with chronic conditions. To date, there have been few studies exploring the online support available for tic disorders, such as Tourette syndrome. An exploratory study looking at users' experiences with using online support communities for tic disorders suggested that members used such communities to share experiences, information, and strategies for tic management.

Objective: To build on these preliminary findings, this study examined the provision of social support in an online community for Tourette syndrome.

Methods: Data were collected from one publicly available online support community for Tourette syndrome and tics, from its inception to December 2019, by randomly selecting 10% of posts and their corresponding comments from each year for analysis. This resulted in 510 unique posts and 3802 comments posted from 1270 unique usernames. The data were analyzed using inductive thematic analysis.

Results: The findings of this study suggest that users utilized the online community as a multifaceted virtual place where they could share and ask for information about tics, unload and share their feelings arising from living with Tourette syndrome, find people facing similar situations and experiences, and freely share the realities of living with Tourette syndrome.

Conclusions: The results complement the findings from a preliminary study and suggest that online support communities have a potentially valuable role as a mechanism for sharing and gaining information on illness experiences from similar peers experiencing tics and can promote self-management of tics. Limitations and recommendations for future research are discussed.

(J Med Internet Res 2022;24(10):e34403) doi:10.2196/34403

KEYWORDS

Tourette syndrome; tic disorders; social support; online support communities; online health communities; thematic analysis; online support; peer support; support group; Tourette; online health community

Introduction

Tic disorders—such as Tourette syndrome and persistent/chronic tic disorder—are noncurable neurodevelopmental conditions characterized by persistent, involuntary outbursts referred to as tics [1]. Tic disorders usually have their onset in childhood: For many, tics tend to decline or disappear in adulthood, while for others, tics persist [2]. Tics are typically rapid and repetitive and can be vocal (ie, sounds) or motor (ie, spasm-like contractions of muscles). For an individual to meet a diagnosis of persistent/chronic tic disorder, 1 or more vocal or motor tics

must be present for more than 1 year since onset; for Tourette syndrome, 2 or more motor tics and at least 1 vocal tic must be present [2]. Tourette syndrome affects approximately 1% of the population worldwide and is more common in men [2,3]. A common public misunderstanding is that coprolalia (tics that involve expressing obscene words) is a key characteristic of Tourette syndrome; however, this affects less than 20% of people with the condition during their lifetime [4] but is independently associated with poorer quality of life, tic severity, and a range of additional neuropsychiatric difficulties [5].

Tic disorders are impairing conditions that often negatively impact an individual's quality of life. Both children and adults with tic disorders frequently report social difficulties, such as being stigmatized or their condition affecting their relationship with family and friends [6]. Negative impact on physical health (eg, pain, injury), decreased self-esteem and mental well-being, functional impairment (eg, delayed progress at school), and decreased overall quality of life have been reported [6-8]. People with tic disorders often experience adverse reactions from other people due to their tics, which include receiving unwanted attention, being excluded, or being bullied [9]. A recent meta-synthesis of 10 studies found that young people often felt ashamed and insecure about their tics, while some adolescents reported defending their rights to be "visible" with Tourette syndrome, and some parents tried to hide or justify their child's tics to strangers, creating tension within the family [10]. In the same review, adults living with Tourette syndrome reported negative experiences across the lifespan, including at school and in workplace environments, and negative experiences with treatment, such as perceiving health professionals as lacking sympathy and knowledge. Furthermore, adults with Tourette syndrome have identified loneliness as a common experience, with readily available support often lacking [11]. Parents of children with Tourette syndrome also frequently report social isolation [12] and have emphasized the significance of using support groups as a "bridge to the outside world" [13].

Social support can be considered as "support accessible to an individual through social ties to other individuals, groups and the larger community" [14]. There are many different functions to social support, including emotional support, informational support, and moral support [15]. It can be a valuable resource when coping with and adjusting to chronic conditions and can promote physical functioning [16] and psychosocial well-being [17,18]. Receiving social support from peers with Tourette syndrome has been reported to help reduce isolation, create feelings of acceptance, and help individuals cope with tics—and can help families too [10-13,19]. Social support may also help counteract the impact of tic disorders upon the psychosocial well-being of individuals and their families [20]. Nevertheless, in-person social support is often unavailable for people with tic disorders due to its rarity [11].

Online support communities (OSCs)—also known by other terms including online support groups, online health communities, and online support forums [21]—have been present for at least 40 years [22], with changes in internet access and increased ownership in internet-enabled technologies over time increasing access to such online spaces. Research into online support groups for health issues emerged in the 1990s, with several reviews published in the early 2000s looking at their usage and effects upon health and social outcomes [23-25]. High-quality trials are lacking in this field; findings from a small number of randomized controlled trials of OSCs have reported mixed results on mental health outcomes [26,27]. However, survey and interview-based research with users of OSCs for many different chronic conditions and health issues (eg, self-harm, insomnia, endometriosis, cancers, polycystic ovary syndrome) have identified the ways in which these virtual communities can be helpful (and unhelpful) to people experiencing a chronic condition or health issue as well as their caregivers [28-34]. Commonly reported benefits of OSCs include having an online space to share and find emotional support for psychological responses to health issues from people who understand it through lived experiences; reducing isolation in facing their health issue by connecting and receiving support from nonjudgmental peers; acquiring new knowledge about their health condition to empower and inform self-management, coping strategies and treatment decisions, and how they interact with health care professionals; and a space where they can create their own narrative and story of their health experiences [28-34]. OSCs may be hosted by an organization (eg, on a health charity's website), on social networking platforms (eg, Facebook, Reddit), or independently of these [22]. These virtual communities can provide an accessible way of gaining social, emotional, and informational support from peers experiencing the same health issue while removing the geographical and temporal barriers to involvement [29,35]. They may be particularly useful for individuals with rare neurodevelopmental conditions, since they provide opportunities to connect with a larger network of peers than they could find locally. Studies have reported that OSC participation has a positive influence on social engagement in children and young people with neurodevelopmental conditions and has potential to facilitate social networking and support [36].

Several studies have examined the types of social support being exchanged within such online communities serving a range of health conditions [19,37-45]. For instance, Coulson and colleagues [39] studied the provision of social support in messages posted to a Huntington disease online community and reported that group members frequently offered informational (advice, referral, teaching, situation appraisal), emotional (relationship, confidentiality, affection, sympathy, validation/empathy, encouragement, prayer, relief of blame), network (access, presence, companionship, express willingness), and esteem (compliment, anchorage) support, while comparatively few offered tangible assistance (perform direct task, active participation). Attard and Coulson [38] looked at the experiences of members of Parkinson disease OSCs and found that participation allowed members to share experiences and knowledge and to develop friendships, as well as helping them cope with the challenges of living with the condition.

Nevertheless, participation in an OSC is not always a positive experience. Compared with identifying the positives, findings are less consistent for disadvantages of online support. Lack of replies may lead members to feel rejected, members may be rude or judgmental to others due to the greater anonymity, misunderstandings are common due to the format of

communication, and information overload may be experienced [38,46]. Furthermore, there is a potential for inaccurate or harmful information to be shared, members who share their complications from treatment can cause others to feel anxious, and success stories may inspire jealousy or hostile interactions within the community [46].

To date, there has been little attention devoted to understanding the role of OSCs for tic disorders. One recent exploratory online survey suggested that online support could "bridge the gap" in accessing support across the course of tics and was a platform through which members could share experiences, information, and strategies for tic management [47]. It can be difficult for patients with tic disorders to access specialist tic services for many environmental and systemic reasons, including issues with health care funding, service delays, referral issues, and a lack of trained specialists [48,49]. This means many people with tics and their families are managing this chronic condition independently—with online support groups helping to "fill this gap" in accessing specialist help. Accessing support online from similar peers may impact a person's illness experience [38]; for people with tic disorders and their families, members reported that participation influenced their decisions about health care, resulted in improvements in psychological well-being, increased confidence, and resulted in greater acceptance of their tics. [47]. Exploring the types of social support provision and communication online may help in understanding the social support needs specific to people with tic disorders and their families, but, as yet, there has been no attempt to understand how social support is enacted within OSCs for tic disorders.

Several studies have taken deductive approaches to analyze and classify posts in OSCs, for example by applying a social support typology such as the Social Support Behavior Code by Cutrona and Suhr [50]. Taking an inductive approach to analyzing data from naturalistic online communication means researchers can approach data without assumptions about what types of social support they expect to be found, with findings being more reflective of the data set [21]. Furthermore, it is uncertain whether findings from inductive studies map to or are dissimilar to findings from studies using deductive approaches. Therefore, in order to address this, our study sought to inductively analyze the content of messages posted in order to identify and describe the provision of social support within an online community for Tourette syndrome and tics.

Methods

Data Collection

Data were collected from one publicly available asynchronous online community devoted to Tourette syndrome and tics. Within this online community, "threads" are started by an individual creating a "post." Other users and the original poster can reply to a post ("comments"), which in turn can be replied to, giving each thread a complex structure. Posts can be text-based or linked to websites and other media (eg, images, videos). The online community has moderators tasked with moderating the community by removing or locking posts or comments if they break the community's specified rules.

All messages posted since the community's inception up to December 31, 2019, were eligible to be included in the data set. During this time, a total of 5382 threads were initiated within the OSC. Each thread title was inspected, and spam messages (eg, adverts) were removed, leaving a total of 5105 threads. Each thread was assigned a number, and using a random number generator [51], 10% of these threads from each year were randomly sampled for analysis. This random sampling approach was taken to ensure that threads were not subjectively chosen by the authors and to reduce potential bias from sampling within a specific timeframe [21,38]. From the 510 threads downloaded for analysis, there was a total of 4312 individual messages (ie, 510 initial messages with 3802 replies) posted from 1270 unique usernames. The content of the 510 threads was downloaded into Microsoft Word.

Ethical Considerations

The study was granted ethics permission from the Division of Rehabilitation Aging and Wellbeing ethics committee at the University of Nottingham (MEDS4008-20-17). The ownership and use of online community content for research purposes are subject to much debate [21]; for this study, previous studies utilizing similar methodologies were used to guide ethical considerations [19,38,44]. The study adhered to ethical guidelines for internet-mediated research developed by the British Psychological Society [52], including anonymizing the name of the online community. No consent was obtained from users to analyze data, as data were taken from one publicly available online community that did not require registration to read or access posts [38]. Usernames were only used to identify the number of unique users included in analyzed posts and were not used in data analysis. Quotes used to illustrate findings were paraphrased to prevent traceability in online searches [21]. Any potentially identifiable information (eg, names of people, places, health care services) was removed from the data.

Data Analysis

Due to the exploratory nature of the study and the lack of prior research, the data were analyzed using data-driven (inductive) thematic analysis [53]. The first step was getting acquainted with the data: The lead author (MJS) read through the whole data set multiple times. Second, initial short codes reflecting the data were created across the data set using the computer software ATLAS.ti. The third step was generating potential themes by making a list of the initial codes and moving the items around in the list until there were clusters of similar codes. This step was completed combining ATLAS.ti and Post-it notes arranged into sorting piles. These initial themes and subthemes were put in a thematic map and reviewed and refined multiple times. The third author (EBD) reviewed coded data and supported the creation of themes and subthemes. The fourth step was reviewing the themes and ensuring that the themes captured the researcher's impression of the data and were independent from one another. The fifth step was defining and naming the themes and subthemes after systematically reviewing them. As in the study by Meade et al [42], the thematic map was discussed and refined with the third author before a final version was agreed upon.

Results

Our thematic analysis generated 4 main themes: (1) “A place to share and ask for information about tics”; (2) “A place to unload my feelings”; (3) “A place where I can find people like me”; and (4) “A place where I can freely share the realities of living with Tourette's.”

A Place to Share and Ask for Information About Tics

Posts to the online community often consisted of members requesting information and advice. The most common request concerned tic management and triggers of tics. Members frequently asked questions regarding ways to reduce, suppress, or redirect tics (“Does anyone know how to relax facial tics?”); treatment options and their effectiveness and side effects (“I thought about trying to suppress my tics by medication. What are the possibilities? Are there any negative long term affects?”); tic severity and frequency (“Is there any research or information of Tourette’s worsening in adulthood?”); tic triggers (“Can it be that alcohol triggers my tics?”); and suggestibility (“Does reading about other people's tics make yours worse?”). Another common request was for advice regarding the diagnosis of Tourette syndrome. Members often described their symptoms and asked the community whether this was Tourette syndrome or if what they described were actually tics. Community members also asked questions around the benefits or disadvantages of having a Tourette syndrome diagnosis. Information and advice regarding the management of comorbid conditions (most often attention-deficit hyperactivity disorder, obsessive compulsive disorder, depression, and anxiety) were also evident. Some posts came from members who used the community to seek information or advice on ways to support their loved one with the syndrome (“What is the best thing to do for someone who is having an attack?”). Members also requested information and advice regarding a range of Tourette syndrome–related everyday issues, for example, how to talk to loved ones about the condition, employment, discrimination, preventing damage to health from tics, and relationship advice.

In response to requests for information and advice, members often replied with comments giving advice and suggestions. Some community members posted advice or suggestions around what worked for them. Members most frequently shared advice for and experiences with managing tics. Prescription medication (frequently mentioned were muscle relaxants, alpha-agonist hypotensive medications, antipsychotics, and carbonic anhydrase inhibitors) and their benefits and risks were discussed. They also discussed and recommended other treatment options (eg, cannabis, herbal remedies, dietary changes and supplements, stress reduction, massage, music, behavioral therapies for tics) and often would share their methods to replace or suppress tics. Some stated that their tics lessen or completely stop when they do something that they enjoy or something that requires intense concentration (eg, playing a musical instrument). Discussions took place around diagnosis: Members shared their anecdotal experience of getting diagnosed with Tourette syndrome or a tic disorder, advice regarding health care professionals, benefits of a diagnosis (“a diagnosis gives you validation, I didn’t realise how much of a difference it makes!”; “getting diagnosed could save you from a law suit or getting fired”), and reasons for not pursuing a diagnosis (“I feel that my tics aren't a big issue, and I don't want to waste the doctor’s time”). In response to posts in which the member posting requested advice regarding diagnosis or shared unusual symptoms, other members often encouraged the advice seeker to see a health care professional and suggested caution (“We are not neurologists”) when giving advice or information. Community members also shared factual information about Tourette syndrome, tics, and comorbid conditions (eg, diagnostic criteria). They gave advice to loved ones of people with the condition on how to best support them, for example, communicate openly, not reacting to tics, understand and learn more about it, listen, reassure them, and connect them to the Tourette syndrome community. Members also offered advice regarding legal protection of Tourette syndrome as a disability and suggestions on dealing with discrimination (“they have to reasonably accommodate it and can't fire you for it, under the disability law”). They also signposted members to other sources of help or information related to the condition (eg, charities, research articles, health care professionals, other OSCs, TV shows, and films).

A Place to Unload My Feelings

The online community was frequently used to vent emotions, such as expressing emotional pain and frustration resulting from living with Tourette syndrome, often in long posts or comments, and sometimes indicating that the community member needed to “vent/get it off their chest” (“The struggle is real ... ... It’s a nightmare and I needed to vent”). Community members often expressed empathy and understanding when replying to posts in which others had shared a tic or experience (ie, verbal description or picture/meme), stating that they had similar tics or experiences and can relate and understand their pain (“I feel your pain!”; “I definitely understand that feeling”). Members also offered reassurance (“That's all perfectly normal”; “You will be ok”) and validated other members’ emotions and experiences (“That must be hard to deal with”; “You should be mad. This is not ok.”). Kind wishes (“Good luck!”) were numerous, and members also expressed their sympathies in response to posts or comments disclosing negative experiences (“I’m so sorry! That really sucks!”).

Members encouraged each other to accept their tics (“You also don’t have to hide your tics! ...You’re unique in your own way and you shouldn’t have to hide yourself!”; “Stay twitchy my friend”) as well as general support (“I know you can do it”; “Hang in there”). Members also praised each other for achievements and sharing created artworks (“I am so incredibly proud of you!”). Humor appeared to be one way to cope with Tourette syndrome, as some members reported that they joke with people “in the real world” about it (“I try to make fun of it to see the social anxiety about it and usually it gets a few laughs… dark humour to some but I would rather people laugh at the jokes about my Tourette’s”) and recommended this coping strategy to others. Members also requested and shared uplifting or encouraging “success” stories and expressed delight in the success of others.

A Place Where I Can Find People Like Me

Community members expressed gratitude to others for the advice and support offered from other users, who understood the kinds of adversity they faced due to having Tourette syndrome ("I just wanted to say I really REALLY enjoy this place. Seeing other people having to grow up and deal with the same I went through is amazing. I knew no one else, so I could never talk about the issues I was dealing with at school etc."). In these posts and comments, members reached out to members in a variety of ways including offering to connect through private message, sharing personal information, and providing updates to previous posts.

Some negative aspects were also identified. Some members were hostile towards others when they disagreed with posts or comments. Some "gatekeeping" issues were also observed as some members complained about the quantity of posts requesting advice regarding diagnosis ("This community has just become a place to ask if you have Tourette's rather than discussing tics and how to deal with them"). This possibly led one individual to wonder if they were "allowed" in the OSC, as their Tourette syndrome was undiagnosed. Other users argued that even though the frequent diagnosis-related posts were irritating for some users, "the pros of this remaining a safe place for (mostly) young people to come and ask about TS as it possibly relates to them far outweigh the negatives."

A Place Where I Can Freely Share the Realities of Living With Tourette's

Members described or showed their tics or suspected tics. The tics shared included vocal and motor tics, and some members shared what their first tics were and some unusual tics and tic presentations (tics in sleep, paralysis tics, and tic attacks). Some members also shared personal videos of their tics. Members often responded with sharing their own tics, and many commented that seeing or hearing about other members' tics was comforting and made them "feel better." Members sought and offered solidarity ("Please let me know I'm not the only weird one in having this tic?"; "we all do our best to help each other out here and no matter what we're always here if you needs support!"). Premonitory urges and physical issues caused by tics (eg, headache, pain, dental issues, injury) were also discussed.

Members shared their unique realities of living with Tourette syndrome, such as what triggers their tics, how their tics wax and wane, and comorbid conditions with which they live. In some posts, members described how their tics felt like to them, while users expressed their Tourette syndrome experience through art. Members shared their anecdotal experiences of dealing with others' misunderstandings, such as rude remarks, staring, bullying, abuse, accusations of "faking" tics, inadequate Tourette syndrome–related health care, and misrepresentation in the media. The impact of Tourette syndrome on their everyday life was also discussed, for instance, tics limiting people by interfering with schoolwork, everyday activities, social life, or relationships. Members emphasized the importance of in-person social support ("having friends and a partner who doesn't respond negatively to my tics has been very helpful") and shared some positive support experiences ("at uni I was accepted and understood, accommodated for and never talked down to, or ridiculed"). Members shared their concerns about "passing on" Tourette syndrome and the complexities of family planning ("Me and my wife are trying to have kids ...I don't want my kid to have to deal with tics and all of this... It's a constant moral dilemma").

Members discussed whether they disclose their tics to other people ("I always just let my co-workers know so that they don't think I'm on drugs or something"; "don't fancy getting them involved so they don't worry. Or maybe …they'd think less of me."). Members also discussed whether they should embrace or deny their tics ("tics are part of who I am"; "no matter how much I try to accept it, I am unable to"), and some discussed that they are sometimes unsure if they are "faking" having tics or Tourette syndrome ("I'm worried that the doctor will not believe me because even i don't believe myself when i speak about the tics").

Discussion

Principal Findings

To date, few studies have looked at social support within OSCs for individuals living with Tourette syndrome or tic disorders. This study aimed to examine the provision of social support in 1 online Tourette syndrome community through an inductive thematic analysis of postings. The findings of this study suggest that users utilized the online community as a multifaceted virtual place where they could share and ask for information about tics, unload their feelings, find people facing similar situations and experiences, and freely share the realities of living with Tourette syndrome. Compared with studies taking similar inductive approaches to analyzing messages in online support groups [19,38,40,41], this study had a large data set reflecting 10% of all threads made in the community since its inception.

The findings in this analysis appear to align with those of previous studies using similar methodologies and finding that OSCs for several chronic conditions tend to report similar functions of social support, including OSCs as a valued virtual place for informational and emotional support from peers experiencing the same health issue, sharing personal experiences with others with empirical lived understanding, and sharing and "venting" emotional reactions typically common with experiencing a chronic condition [19,38,45]. The results from this study complement the findings from a previous online survey exploring users' experiences with participating in Tourette syndrome/tic disorder OSCs [47]. What this study adds is that, by analyzing posts created naturalistically over time in 1 group, it has identified additional social support needs unique to tic disorders. For example, the "A place to unload my feelings" theme identified an emotional support need relating to users' discussions that they may be "faking" their tics or symptoms and dealing with victimization from other people. This may be unique to people with tics, given the nature of tics and the socially stigmatizing nature of tic disorders. These findings may also be of value to health care professionals in further understanding the emotional needs of patients with tic disorders and how they can be supported.

In line with the findings by Meade et al [42], the results of this study suggest that OSCs have a potentially valuable role as a mechanism for sharing and gaining information on illness experiences and empowering individuals and supportive others in relation to self-management of neurodevelopmental conditions such as Tourette syndrome. The online nature of the support community may also aid social support by providing an anonymous environment through which users can disclose information that they would find difficult to express in person [39] or discuss sensitive topics [28]. For people with Tourette syndrome, online communities may provide an accessible and inclusive space, where they can gain social support not easily available in their offline worlds—and from peers who understand what it is like to be socially excluded and "different" due to their tics [47]. Regular online community users may also gain additional benefits as a consequence of providing support to other users, in accordance with the helper-therapy principle [54], which suggests that people also help themselves when helping others, by taking on important social roles, developing their coping skills, and directing their focus away from their own problems.

Limitations and Future Directions

Despite the insights generated through this study, there are limitations that should be considered. First, our analysis focused on textual data; therefore, conversations between community members did not have any associated nonverbal conversational cues. This arguably makes our task more difficult and creates risks around misinterpretation of the data. These risks are exacerbated by the fact we did not engage community members in the analytical process. However, to mitigate against this risk, each post analyzed was done so in the context of the full conversation thread. Second, this study looked at only 1 Tourette syndrome OSC; therefore, the extent to which the results can be generalized may be limited, as other communities may differ in their structure, membership, community dynamics, and types of social support requested or offered. Third, little is known about the demographic characteristics of the community members included in this study; therefore, it is difficult to assess how representative they are of the wider population of individuals with Tourette syndrome or tic disorders. Additionally, users of the online community were a mix of people with a tic disorder themselves and supportive others (eg, parents, caregivers, partners). It was not always obvious from the posts and threads whether the user was a person with tic disorder themselves or a caregiver or supportive other of someone with tics; therefore, it was not possible to do a subanalysis by type of user. Although not a limitation in itself, there may be some similarities and differences in the benefits and kinds of social support these 2 groups seek online, given their different roles in the illness experience.

Finally, this analysis was conducted on data collected prior to the COVID-19 pandemic. During the pandemic, there have been reports from clinicians regarding increased numbers of referrals to specialist tic clinics or services—particularly from adolescent girls—with some clinicians suggesting this new influx of patients is linked to online video-based media from Tourette syndrome content creators (eg, TikTok, YouTube) [55,56]. Understandably, given this and societal changes during the pandemic (eg, potentially not being able to access usual health care systems or peer support), usage of online communities for Tourette syndrome and tic disorders may have changed or membership increased substantially during this time—and this was not captured in our analysis. Looking at publicly available statistics for the 1 OSC used in this study, as of January 1, 2022, there was a 235% increase in registered users over 2 years. Between its inception and December 31, 2019, there were 5382 threads posted to the OSC; between January 1, 2020, and January 1, 2022, there was a 125% increase in threads posted over 2 years. These data suggest increased membership and usage over the past 2 years, which overlaps with the COVID-19 pandemic.

Given the present study only looked at 1 online support group, future research may wish to explore the social support provided across multiple OSCs for Tourette syndrome and tic disorders. Studies could also explore the online experiences of individuals with Tourette syndrome or tic disorders and caregivers separately, as previous research suggests potential mismatch between these 2 groups of users [47]. Furthermore, this study analyzed the content of online communication, which is arguably in the public domain. It is unknown whether discourse varies in private OSCs, as well as communities developed using a range of different platforms and modalities of communication (eg, Facebook, Discord), and this could potentially be explored in further research.

Conclusion

Online support may be a useful, easily accessible addition to traditional forms of support for people with Tourette syndrome or tic disorders and their supportive others, where they can share and request information about tics, unload their feelings, find people with similar experiences, and share the realities of living with the condition.

Acknowledgments

This work was cofunded by the National Institute for Health and Care Research (NIHR) MindTech Medtech Cooperative and the NIHR Nottingham Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the National Health Service, NIHR, or Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. Neurodevelopmental Disorders (5th Edition). The Diagnostic and Statistical Manual of Mental Disorders. Arlington, VA: American Psychiatric Association; 2013.
2. Set K, Warner J. Tourette syndrome in children: An update. Curr Probl Pediatr Adolesc Health Care 2021 Jul;51(7):101032 [FREE Full text] [doi: 10.1016/j.cppeds.2021.101032] [Medline: 34305006]
3. Robertson MM, Eapen V, Cavanna AE. The international prevalence, epidemiology, and clinical phenomenology of Tourette syndrome: a cross-cultural perspective. J Psychosom Res 2009 Dec;67(6):475-483. [doi: 10.1016/j.jpsychores.2009.07.010] [Medline: 19913651]
4. Freeman R, Zinner S, Müller-Vahl KR, Fast D, Burd L, Kano Y, et al. Coprophenomena in Tourette syndrome. Dev Med Child Neurol 2009 Mar;51(3):218-227 [FREE Full text] [doi: 10.1111/j.1469-8749.2008.03135.x] [Medline: 19183216]
5. Eddy CM, Cavanna AE. On being your own worst enemy: an investigation of socially inappropriate symptoms in Tourette syndrome. J Psychiatr Res 2013 Sep;47(9):1259-1263. [doi: 10.1016/j.jpsychires.2013.05.019] [Medline: 23768869]
6. Eapen V, Cavanna AE, Robertson MM. Comorbidities, social impact, and quality of life in Tourette syndrome. Front Psychiatry 2016 Jun 06;7:97 [FREE Full text] [doi: 10.3389/fpsyt.2016.00097] [Medline: 27375503]
7. Wadman R, Glazebrook C, Beer C, Jackson GM. Difficulties experienced by young people with Tourette syndrome in secondary school: a mixed methods description of self, parent and staff perspectives. BMC Psychiatry 2016 Jan 20;16(1):14 [FREE Full text] [doi: 10.1186/s12888-016-0717-9] [Medline: 26792211]
8. Conelea CA, Woods DW, Zinner SH, Budman CL, Murphy TK, Scahill LD, et al. The impact of Tourette syndrome in adults: results from the Tourette syndrome impact survey. Community Ment Health J 2013 Feb;49(1):110-120. [doi: 10.1007/s10597-011-9465-y] [Medline: 22052430]
9. Edwards KR, Mendlowitz S, Jackson E, Champigny C, Specht M, Arnold P, et al. A qualitative exploration of the experiences of children and adolescents with Tourette syndrome. J Can Acad Child Adolesc Psychiatry 2017;26(1):39-44 [FREE Full text] [Medline: 28331502]
10. Smith H, Fox JRE, Trayner P. The lived experiences of individuals with Tourette syndrome or tic disorders: a meta-synthesis of qualitative studies. Br J Psychol 2015 Nov;106(4):609-634. [doi: 10.1111/bjop.12118] [Medline: 25721405]
11. Malli MA, Forrester-Jones R, Triantafyllopoulou P. "Tourette's Is a Lonely Place": an interpretative phenomenological analysis of the personal experience and identity of adults with Tourette's syndrome. J Dev Phys Disabil 2019 May 14;31(6):819-845. [doi: 10.1007/s10882-019-09676-2]
12. Ludlow AK, Brown R, Schulz J. A qualitative exploration of the daily experiences and challenges faced by parents and caregivers of children with Tourette's syndrome. J Health Psychol 2018 Dec;23(14):1790-1799. [doi: 10.1177/1359105316669878] [Medline: 27682339]
13. O'Hare D, Eapen V, Grove R, Helmes E, Mcbain K, Reece J. Youth with Tourette syndrome: Parental perceptions and experiences in the Australian context. Australian Journal of Psychology 2020 Nov 20;69(1):48-57. [doi: 10.1111/ajpy.12111]
14. Lin N, Simeone RS, Ensel WM, Kuo W. Social support, stressful life events, and illness: a model and an empirical test. J Health Soc Behav 1979 Jun;20(2):108-119. [Medline: 479524]
15. Costa-Cordella S, Arevalo-Romero C, Parada FJ, Rossi A. Social support and cognition: a systematic review. Front Psychol 2021 Feb 23;12:637060 [FREE Full text] [doi: 10.3389/fpsyg.2021.637060] [Medline: 33708164]
16. Rutledge T, Reis SE, Olson M, Owens J, Kelsey SF, Pepine CJ, National Heart, Lung, and Blood Institute. Social networks are associated with lower mortality rates among women with suspected coronary disease: the National Heart, Lung, and Blood Institute-Sponsored Women's Ischemia Syndrome Evaluation study. Psychosom Med 2004;66(6):882-888. [doi: 10.1097/01.psy.0000145819.94041.52] [Medline: 15564353]
17. Turner-Cobb JM, Gore-Felton C, Marouf F, Koopman C, Kim P, Israelski D, et al. Coping, social support, and attachment style as psychosocial correlates of adjustment in men and women with HIV/AIDS. J Behav Med 2002 Aug;25(4):337-353. [doi: 10.1023/a:1015814431481] [Medline: 12136496]
18. Smith CE, Fernengel K, Holcroft C, Gerald K, Marien L. Meta-analysis of the associations between social support and health outcomes. Ann Behav Med 1994;16(4):352-362 [FREE Full text] [doi: 10.1093/abm/16.4.352]
19. Ainscough E, Smith SN, Greenwell K, Hoare DJ. Findings and ethical considerations from a thematic analysis of threads within tinnitus online support groups. Am J Audiol 2018 Nov 19;27(3S):503-512 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0013] [Medline: 30452754]
20. Goussé V, Czernecki V, Denis P, Stilgenbauer J, Deniau E, Hartmann A. Impact of perceived stress, anxiety-depression and social support on coping strategies of parents having a child with Gilles de la Tourette syndrome. Arch Psychiatr Nurs 2016 Feb;30(1):109-113. [doi: 10.1016/j.apnu.2015.08.017] [Medline: 26804511]
21. Smedley RM, Coulson NS. A practical guide to analysing online support forums. Qualitative Research in Psychology 2018 May 30;18(1):76-103. [doi: 10.1080/14780887.2018.1475532]
22. Potts HWW. Online support groups: An overlooked resource for patients. Health Information on the Internet 2005;44(1):6-8 [FREE Full text]
23. Valaitis R, Ciliska D. Electronic social support groups to improve health. Effective Public Health Practice Project (EPHPP). 2000. URL: https://old.hamilton.ca/phcs/ephpp/Research/Full-Reviews/ESSG-Review.PDF [accessed 2022-09-14]

24. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
25. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003;21(3):136-142. [doi: 10.1097/00024665-200305000-00010] [Medline: 12792194]
26. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015 May 19;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
27. Griffiths KM, Calear AL, Banfield M. Systematic review on internet support groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
28. Buchanan H, Coulson NS. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269. [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]
29. Shoebotham A, Coulson NS. Therapeutic affordances of online support group use in women with endometriosis. J Med Internet Res 2016 May 09;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]
30. Huber J, Maatz P, Muck T, Keck B, Friederich H, Herzog W, et al. The effect of an online support group on patients treatment decisions for localized prostate cancer: An online survey. Urol Oncol 2017 Feb;35(2):37.e19-37.e28 [FREE Full text] [doi: 10.1016/j.urolonc.2016.09.010] [Medline: 27810256]
31. Holbrey S, Coulson NS. A qualitative investigation of the impact of peer to peer online support for women living with polycystic ovary syndrome. BMC Womens Health 2013 Dec 17;13(1):51 [FREE Full text] [doi: 10.1186/1472-6874-13-51] [Medline: 24341398]
32. Coulson NS, Bullock E, Rodham K. Exploring the therapeutic affordances of self-harm online support communities: an online survey of members. JMIR Ment Health 2017 Oct 13;4(4):e44 [FREE Full text] [doi: 10.2196/mental.8084] [Medline: 29030324]
33. Coulson NS, Smedley R, Bostock S, Kyle SD, Gollancz R, Luik AI, et al. The pros and cons of getting engaged in an online social community embedded within digital cognitive behavioral therapy for insomnia: survey among users. J Med Internet Res 2016 Apr 25;18(4):e88 [FREE Full text] [doi: 10.2196/jmir.5654] [Medline: 27113540]
34. McKechnie V, Barker C, Stott J. The effectiveness of an Internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014 Feb 28;16(2):e68 [FREE Full text] [doi: 10.2196/jmir.3166] [Medline: 24583789]
35. Coulson NS, Knibb RC. Coping with food allergy: exploring the role of the online support group. Cyberpsychol Behav 2007 Feb;10(1):145-148. [doi: 10.1089/cpb.2006.9978] [Medline: 17305463]
36. Saxena S, Mitchell J, Ehsan A, Majnemer A, Shikako-Thomas K. Online peer mentorship programmes for children and adolescents with neurodevelopmental disabilities: A systematic review. Child Care Health Dev 2020 Jan;46(1):132-148. [doi: 10.1111/cch.12726] [Medline: 31782542]
37. Anderson S, Clarke V. Disgust, shame and the psychosocial impact of skin picking: Evidence from an online support forum. J Health Psychol 2019 Nov 10;24(13):1773-1784. [doi: 10.1177/1359105317700254] [Medline: 28810443]
38. Attard A, Coulson N. A thematic analysis of patient communication in Parkinson's disease online support group discussion forums. Comput Human Behav 2012 Mar;28(2):500-506 [FREE Full text] [doi: 10.1016/j.chb.2011.10.022]
39. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
40. Hanna E, Gough B. Searching for help online: An analysis of peer-to-peer posts on a male-only infertility forum. J Health Psychol 2018 Jun;23(7):917-928. [doi: 10.1177/1359105316644038] [Medline: 27106090]
41. Malik SH, Coulson N. The male experience of infertility: a thematic analysis of an online infertility support group bulletin board. J Reprod Infant Psychol 2008 Feb;26(1):18-30. [doi: 10.1080/02646830701759777]
42. Meade O, Buchanan H, Coulson N. The use of an online support group for neuromuscular disorders: a thematic analysis of message postings. Disabil Rehabil 2018 Sep 08;40(19):2300-2310. [doi: 10.1080/09638288.2017.1334239] [Medline: 28592203]
43. Morris MT, Daluiski A, Dy CJ. A thematic analysis of online discussion boards for brachial plexus injury. J Hand Surg Am 2016 Aug;41(8):813-818. [doi: 10.1016/j.jhsa.2016.05.007] [Medline: 27311862]
44. Smedley RM, Coulson NS. Genetic testing for Huntington's disease: A thematic analysis of online support community messages. J Health Psychol 2021 Mar;26(4):580-594. [doi: 10.1177/1359105319826340] [Medline: 30696276]
45. Thoren EM, Metze B, Bührer C, Garten L. Online support for parents of preterm infants: a qualitative and content analysis of Facebook 'preemie' groups. Arch Dis Child Fetal Neonatal Ed 2013 Nov;98(6):F534-F538. [doi: 10.1136/archdischild-2012-303572] [Medline: 23759520]
46. Turner JA. Online support groups: the good, the bad, and the motivated. J Consum Health Internet 2017 Mar 10;21(1):11-25. [doi: 10.1080/15398285.2017.1279930]

47. Perkins V, Coulson NS, Davies EB. Using online support communities for Tourette syndrome and tic disorders: online survey of users' experiences. J Med Internet Res 2020 Nov 03;22(11):e18099 [FREE Full text] [doi: 10.2196/18099] [Medline: 33141089]
48. Cuenca J, Glazebrook C, Kendall T, Hedderly T, Heyman I, Jackson G, et al. Perceptions of treatment for tics among young people with Tourette syndrome and their parents: a mixed methods study. BMC Psychiatry 2015 Mar 11;15:46 [FREE Full text] [doi: 10.1186/s12888-015-0430-0] [Medline: 25879205]
49. Woods DW, Conelea CA, Himle MB. Behavior therapy for Tourette's disorder: Utilization in a community sample and an emerging area of practice for psychologists. Prof Psychol: Research and Practice 2010 Dec;41(6):518-525. [doi: 10.1037/a0021709]
50. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 2016 Jun 30;19(2):154-174. [doi: 10.1177/009365092019002002]
51. Random.org. URL: https://www.random.org/ [accessed 2022-09-14]
52. Ethics guidelines for internet-mediated research. British Psychological Society. 2017. URL: https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research [accessed 2022-09-14]
53. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
54. Riessman F. The “helper” therapy principle. Social Work 1965;10(2):27-32. [doi: 10.1093/sw/10.2.27]
55. Heyman I, Liang H, Hedderly T. COVID-19 related increase in childhood tics and tic-like attacks. Arch Dis Child 2021 Mar 06;106(5):420-421. [doi: 10.1136/archdischild-2021-321748] [Medline: 33677431]
56. Pringsheim T, Ganos C, McGuire J, Hedderly T, Woods D, Gilbert D, et al. Rapid onset functional tic-Like behaviors in young females during the COVID-19 pandemic. Mov Disord 2021 Dec;36(12):2707-2713 [FREE Full text] [doi: 10.1002/mds.28778] [Medline: 34387394]

Abbreviations

NIHR: National Institute for Health and Care Research
OSC: online support community

Edited by R Kukafka; submitted 26.10.21; peer-reviewed by J Smith-Merry, H McCall; comments to author 05.12.21; revised version received 15.02.22; accepted 13.03.22; published 04.10.22.

Please cite as:
Soós MJ, Coulson NS, Davies EB
Exploring Social Support in an Online Support Community for Tourette Syndrome and Tic Disorders: Analysis of Postings
J Med Internet Res 2022;24(10):e34403
URL: https://www.jmir.org/2022/10/e34403
doi:10.2196/34403
PMID:36194454

Original Paper

The Use of Close Friends on Instagram, Help-Seeking Willingness, and Suicidality Among Hong Kong Youth: Exploratory Sequential Mixed Methods Study

Sikky Shiqi Chen[1], BSocSci; Tai Pong Lam[1], MD, PhD; Kwok Fai Lam[2,3], PhD; Tak Lam Lo[4], MBBS; David Vai Kiong Chao[5], MBChB; Ki Yan Mak[1], MD; Edmund Wing Wo Lam[1], MBBS; Wai Sin Tang[1], MA; Hoi Yan Chan[1], MSc; Paul Siu Fai Yip[6,7], PhD

[1]Department of Family Medicine and Primary Care, The University of Hong Kong, Hong Kong, China
[2]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
[3]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore
[4]Pamela Youde Nethersole Eastern Hospital, Hong Kong, China
[5]Department of Family Medicine and Primary Health Care, Hospital Authority Kowloon East Cluster, Hong Kong, China
[6]Hong Kong Jockey Club Centre for Suicide Research and Prevention, The University of Hong Kong, Hong Kong, China
[7]Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong, China

Corresponding Author:
Tai Pong Lam, MD, PhD
Department of Family Medicine and Primary Care
The University of Hong Kong
3/F, Ap Lei Chau Clinic, 161 Main Street
Ap Lei Chau
Hong Kong
China
Phone: 852 25185688
Email: tplam@hku.hk

Abstract

Background: Social networking sites (SNSs) have gained popularity in recent years for help seeking and self-distress expression among adolescents. Although online suicidal expression is believed to have major benefits, various concerns have also been raised, particularly around privacy issues. Understanding youths' help-seeking behavior on SNSs is critical for effective suicide prevention; however, most research neglects the impacts of the private SNS context.

Objective: This study aims to examine youths' private SNS use via the new Instagram feature, *Close Friends*, and its association with both online and offline help-seeking willingness as well as youths' suicidality.

Methods: This study employed an exploratory sequential mixed methods approach with a combination of explorative qualitative interviews and a systematic quantitative survey, targeting youth aged 15-19 years in Hong Kong. The motivations for utilizing Close Friends and concerns regarding online expression were addressed in the focus group and individual interviews (n=40). A cross-sectional survey (n=1676) was conducted subsequently with eligible secondary school students to examine the prevalence of Close Friends usage, their online and offline help-seeking willingness, and suicide-related experiences.

Results: A total of 3 primary motives for using Close Friends were identified during interviews, including (1) interaction and help seeking, (2) release of negative emotions, and (3) ventilation and self-expression. Most participants also highlighted the privacy concerns associated with public online communication and the importance of contacting close friends for emotional support. Survey results showed that use of Close Friends was quite prevalent among adolescents (1163/1646, 70.66%), with around 46% (754/1646, 45.81%) of respondents being frequent users. Differences by gender and school academic banding were also revealed. Regarding help-seeking intentions, youths were generally positive about seeking help from peers and friends offline (1010/1266, 79.78%) yet negative about seeking assistance from online friends or professionals with whom they had not yet developed a real-world connection (173/1266, 13.67%). Most notably, frequencies of Close Friends usage were differentially associated with online and offline help-seeking willingness and youths' suicidality. Compared with nonusers, those who had ever used the feature were more likely to seek offline support (adjusted odds ratios [AORs] 1.82-2.36), whereas heavy use of Close

Friends was associated with increased odds of online help-seeking willingness (AOR 1.76, 95% CI 1.06-2.93) and a higher risk of suicidality (AOR 1.53, 95% CI 1.01-2.31).

Conclusions: The popularity of Close Friends reflects the increasing need for private online expression among youth. This study demonstrates the importance of Close Friends for self-expression and private conversation and inadequacy of peer support for suicidal adolescents. Further research is needed to identify the causal relationship between Close Friends usage and help-seeking willingness to guide the advancement of suicide prevention strategies. Researchers and social media platforms may cooperate to co-design a risk monitoring system tailored to the private SNS context, assisting professionals in identifying youth at risk of suicide.

(J Med Internet Res 2022;24(10):e37695) doi:10.2196/37695

KEYWORDS

Close Friends; private online expression; help-seeking willingness; suicide; youth

Introduction

Youths' Online Expression

Online social networking sites (SNSs) are gaining increasing popularity among adolescents in recent years, changing the nature of communication. Instagram is one of the most popular SNSs worldwide [1] and the third most popular SNS in Hong Kong among people aged 16-64 years [2]. It has roughly 1 billion monthly active users [1], with around 40% of them aged 13-24 [3]. Prior studies have shown that SNSs, especially Instagram, have become a favored alternative for at-risk youth to seek help and express their distress [4,5]. Indeed, online expression pertinent to suicidality is thought to have several major benefits, including mitigated social isolation, recovery-oriented encouragement, and alleviation of acute self-harm urges [6,7]. However, in light of concerns around privacy issues, unsympathetic responses from anonymous viewers [8], and the "positivity bias," negative disclosure on SNS is often deemed risky and thus less commonly shared publicly [9]. For example, previous research on Facebook demonstrated that most negative expressions were published exclusively on private pages as opposed to publicly [10].

To date, there has been only minimal discussion about private SNS usage among users with varying levels of mental health issues. It, therefore, remains uncertain as to how private and public online expressions differ in terms of social support seeking. This is the first study to adopt an exploratory sequential mixed methods approach to assess how private online expression is correlated with help-seeking willingness and suicidality among adolescents via *Close Friends* posts on Instagram. It was anticipated that our findings would inform the development of more effective and pragmatic suicide prevention and intervention programs delivered via SNS.

Close Friends: A New Feature for Private Online Expression

Private online interaction has gradually developed with the advancement of technology and now varies in terms of intimacy and extent of self-disclosure [11]. Private expression on SNSs, also known as active private SNS use, is characterized by direct users' interactions in a private setting, generally via instant messaging and personal chatting [12]. To compare the 2 modes of SNS use, public online communication allows large-scale interactions between individuals who have never met in person, whereas private online communication usually occurs in a smaller group and involves friends who have established trust and fundamental mutual understanding offline [13,14].

Aversion to publishing unpleasant content, especially mental health–related disclosures, has been observed on Instagram. On account of context collapse, the most sensitive (such as families and romantic partners) and unintended audiences can browse youths' SNS profiles and navigate their posts [15,16]. Adolescents have shown concerns that posting depressive messages in the public space will be criticized as being offensive by the broad, diverse, or unknown audience [17]. Therefore, some create a fake Instagram account (dubbed Finsta) specifically for negative online expression [18]. On Finsta, users only follow their intimate friends so they are less worried about revealing negativity and vulnerability, and are instead prompted to present authentic and unfiltered self-expression [19]. The emerging demand for private online expression to discuss sensitive issues is reflected in the trend of multiple accounts. The traditional features of SNSs, by contrast, are deficient in terms of private expression, mostly through dyadic messages [20], making it difficult for adolescents to engage with multiple users simultaneously. The format of text-based communication is also restricted in conveying emotions precisely.

Hence, Instagram launched a new feature named Close Friends in late 2018 [21]. Only those added to the list are permitted to view private *Stories* posted in Close Friends. Followers will also be notified that they have been included in one's Close Friends list on Instagram when they see their posts with the special icon. Compared with traditional private communication, Close Friends provides more opportunities for online expression by virtue of its unique features [22], which include the following:

- Close Friends enables parallel multiple interactions in a private context.
- Users have their own absolute control over the members' list, which means they are able to block or remove anyone on the list at any time without notifying them.
- Posts in Close Friends do not predetermine any specific audience.
- Users are able to seek attention via posts with sensitive information, thus help seeking requires less proactivity.

- Users' posts in Close Friends are actively selected and "pressed" to view by the audience who care about their lives.

The aforementioned properties of Close Friends distinguish its content display from those of the conventional tools for private online communication. Indeed, Close Friends is a more visually, graphically, and multimedia-oriented feature compared with instant messaging, which largely relies on text- or voice-based communications. Users of Close Friends are also able to publish posts in vivid formats, such as photos and videos, to capture, share, and showcase personal life moments [23]. Furthermore, the text-based postings in Close Friends are more aesthetically appealing than those in messages or feeds due to the availability of various fonts, animated graphics interchange format, stickers, and predesigned layouts.

Private Online Expression, Help-Seeking Willingness, and Suicidality

Suicide is a major public health concern worldwide, particularly among youth. According to the World Health Organization, suicide is the fourth leading cause of death among those aged 15-19 [24,25]. In Hong Kong, youth aged between 15 and 24 had a suicidal rate below 10 per 100,000 [26-28], which is lower than that of their counterparts, including in European countries and most East Asian neighbors [29]. However, a rising trend in the overall suicide rate among Hong Kong adolescents has been noted throughout the last 2 decades, with rates increasing from 7.7 per 100,000 in 2000 to 9.5 per 100,000 in 2018 [30].

According to the Uses and Gratifications Theory, adolescents are proactive and goal-oriented SNS users who consider their interests and expectations while choosing and using platforms [31]. The need for satisfying diverse motives would result in different online behaviors. Hence, it is essential to acquire a comprehensive understanding of youths' purposes of using the internet. Previous research has investigated the motivations for adolescents' use of Instagram. One such study identified 5 motives, including social engagement, archiving, self-expression, escapism, and peeking [32], while another study recognized 4 major purposes, comprising surveillance, documentation, coolness, and creativity [33]. In addition, our earlier study examined the principal motivations of online expression among Hong Kong youth, including self-expression, emotional ventilation, life sharing and documentation, social interaction, attention seeking, and help seeking [34]. We also identified a positive association between willingness of online help seeking and the motivations of expressing emotions and opinions.

The significance of professional and nonprofessional support has been validated across different populations [35,36], and therefore, most suicide interventions encourage people to seek help. Nevertheless, evidence from previous research suggested mixed findings when it came to the relationship between suicidality and online help seeking among adolescents [37]. For example, several studies found that youth who sought help online were more likely to use the internet for suicide-related purposes [38] and experience social anxiety, psychological distress, self-harm behaviors, and suicide [39,40], while other studies reported that online communication might provide social and emotional support, which could facilitate ones' coping with depression and stress [41-43]. In general, as a large number of studies have indicated, seeking help from peers and friends in real life is preferred by the young population, compared with formal help sources (ie, professionals) and unfamiliar people online [5,44-46]. Adolescents at lower risk of suicide and mental health problems are more likely to engage in offline help seeking from peers and friends [43,47]. Among those who prefer seeking help online, one of the main motivations is to compensate for any deficits in offline support [48,49].

To date, little attention has been paid to private SNS disclosure, particularly among adolescents. Indeed, much uncertainty still exists regarding the relationship between private online expression and suicidality. Only one very recent study of university students found that active private SNS usage was associated with a lower level of suicidality [50], whereas the frequency for each type of SNS usage was not explored. Results of a longitudinal study revealed that both heavy and suicide-related internet use were strongly associated with suicidal ideation (SI) and behaviors [51]. The subgroup with high SNS usage reported more psychiatric problems and social dysfunction as well as limited family or friend support. In terms of private SNS usage and help-seeking intentions, despite a positive association between active private Facebook usage and perceived online social support demonstrated in a youth sample [12], to our knowledge, no study has directly assessed the effect of private online expression on willingness of help seeking, either online or offline.

Rationale for This Study

At present, very little is known about the role Close Friends plays in both online expression and help seeking for suicidality. This study is part of a larger project examining youth suicide in Hong Kong with the specific aim of exploring the use of Close Friends among adolescents. An exploratory sequential mixed methods design was adopted. Qualitative data were first collected through interviews, and major themes were generated and used to facilitate the development of the quantitative instrument. The study objectives are as follows: (1) to determine the prevalence of Close Friends among youth; (2) to ascertain the frequency, purposes, and reasons for using Close Friends; and (3) to investigate its relationship with willingness to seek help both online (from friends and professionals) and offline (from peers and friends); and (4) to examine the association between the use of Close Friends and suicidality.

Methods

Qualitative Approach

Overview

Because of the lack of prior research on Close Friends, qualitative interviews enabled us to obtain a better grasp of how it was used and viewed by youth. Considering the sensitivity of suicide-related topics, most participants first engaged in 1 of the 6 focus group discussions (3-7 persons per session), with those in each group being acquainted with each other. We then conducted 12 in-depth semistructured individual interviews for those who were unable to join focus groups or who had

difficulties elaborating their stories in group interviews on account of the setting or time constraints.

Participant Recruitment

In total, 40 participants aged 15-18 (mean 16.3, SD 1.1) years were recruited, including 31 focus group participants (12 males and 19 females) and 12 individual interview participants (4 males and 8 females); 3 female participants took part in both an individual interview and a focus group discussion. Of the 40 participants, 35 had lived experience of suicide (ie, incidence of SI, self-harm, suicide attempts [SAs], or having helped someone in crisis). The target group for interviews was Hong Kong youth aged 15-19 years with emotional distress or suicidal concern; additionally, those with comparatively fewer problems, but who showed willingness to discuss the topic of suicide, were invited. Purposive sampling was adopted to maximize variation and assure diversity of participants' sociodemographic characteristics (eg, social backgrounds: secondary school students, university freshmen, and school dropouts), clinical and mental health status, and suicide-related experiences. Adolescents who satisfied the recruitment criteria were asked if they were willing to attend the interviews. A major barrier to recruitment was building a rapport with youth with suicidality and earning their trust, as most tended to conceal their past out of fear of the stigma associated with suicide. To facilitate the recruitment process, we solicited recommendations from people who served, or who were closely bonded with this specific group of adolescents, including teachers, medical practitioners, and school social workers, to recommend suitable participants. The circumstances of vulnerable individuals with SA experiences, or prior diagnosis of mental illnesses, were evaluated by 2 mental health professionals to affirm their eligibility for the interviews. Invitation letters and consent forms were sent to eligible participants or, if they were under the age of 18 years, to their parents and guardians.

Procedure and Analysis

Considering the sensitive nature of the suicide topic, interviews were conducted in a face-to-face manner and in a natural, private, and secure environment. Open-ended questions regarding the experiences and motivations of using Close Friends, concerns about online expression, and willingness of help seeking were addressed in the interviews. All the interviews were conducted between September 2018 and November 2019. Each focus group lasted for 1-1.5 hours, and each individual interview lasted around 1 hour. Two experienced facilitators led the interviews in Cantonese (the local dialect) and took charge of data analysis to ensure data trustworthiness.

Reflexive thematic analysis was performed using an inductive approach [52,53]. We focused on both semantic and latent meanings, and adhered to the 6-step framework (Textbox 1 [54]) outlined by Braun and Clarke [54].

Textbox 1. The 6-step framework outlined by Braun and Clarke.

1. Familiarization

- All interviews were audio recorded and transcribed verbatim. Two coders (SSC and HYC) listened to the recordings carefully, read the transcripts iteratively, shared first impressions, and took brief notes.

2. Generating codes

- The cleaned transcripts were entered into the NVivo database (version 12; QSR International Pty Ltd). Initially, both coders independently coded the data. Relevant, informative, and potentially interesting items were encoded with concise and clear codes. SSC and HYC then contrasted the 2 sets of codes and examined how different codes could work together across the data set.

3. Generating initial themes

- Upon coding completion, clustering or splitting of the valid codes was determined based on their patterns, and overarching themes were generated.

4. Developing and revising themes

- All the initial themes were further developed. Through team discussion with 2 senior qualitative researchers (TPL and WST), SSC and HYC revised the overlaps and divergences identified in the initial themes. SSC assessed the coherence of codes within each candidate theme and reread the transcripts to evaluate the congruence between themes and the overall data set.

5. Refining, defining, and naming final themes

- Feedback from the team discussion guided theme refinement. Some candidate themes were combined, split, or discarded before SSC decided and named the final themes. Related quotes under each final theme were collated.

6. Reporting

- The qualitative findings were reported by the research team and critically reviewed by all authors. Principal themes and subthemes were converted into the questionnaire items of the quantitative survey.

Quantitative Approach

Design

A cross-sectional design was adopted in the quantitative phase by the implementation of a self-administered questionnaire survey among secondary students in Hong Kong (aged 15-19 years).

Sample

The target population consisted of students in grades 10-12 who were aged 15-19 years. Invitation letters were sent via postal mail to all the secondary schools in Hong Kong, with 9 schools agreeing to participate. According to the official data released by the Education Bureau of the government in 2019, 150,720 grade 10-12 students were enrolled in 504 local and international secondary schools [55]. According to epidemiological statistics, the prevalence of SI among Hong Kong youth was no more than 25% [56]. On this premise, the required sample size was calculated. Responses from 1801 respondents were expected to have a maximum estimation error (absolute precision) of d=0.02 from the true prevalence rate with a 95% CI.

Data Collection and Questionnaire

Quantitative data were collected between September and October 2019. The questionnaire was anonymous and coded with a reference number (eg, A001) to indicate the schools for data analyses. As required by the ethics committee, we developed a risk management protocol with careful consideration of both the preservation of confidentiality and the facilitation of risk control in schools. Teachers or coordinators at participating schools would be informed of the distribution of suicidal risks among their students. An alert would also be sent out to the school if a certain portion of high-risk cases (ie, ≥25%; referring to youth suicidality rate in Hong Kong) were identified [56].

Questionnaire items were stemmed from qualitative findings, a review of the literature, and feedback from the research team. The final survey consisted of 51 items and took around 15-20 minutes to complete. As part of the larger project, this study contained several sections of the questionnaire, including the frequency (measured on a 4-point Likert scale: 1, never; 2, seldom; 3, sometimes; and 4, often; response options are comparable with those adopted in a previous study [50]) of using Close Friends, willingness of online help seeking, willingness of seeking help from peers and friends, SI and SA experiences in the past 12 months, and sociodemographic information.

To measure respondents' help-seeking willingness, we posed the following question: "When confronted with distressing issues or life difficulties, did you seek help from any of the following in the last 12 months?" The response options were dichotomous (coded as *Yes*=1 or *No*=0) and included both online (online friends and online professionals) and offline (peers, friends, and classmates) resources of help. Respondents' suicide risk was examined by questions on prior suicidal behaviors scored as *Yes*=1 or *No*=0. We asked respondents "Have you considered killing yourself in the past 12 months" and "Have you attempted to kill yourself during the past 12 months?" Nonaffirmative responses to both questions were categorized as "no/low risk" of suicide, affirmative responses to the first question only as "medium risk," and affirmative responses to both questions as "high risk." Investigation on recent SI and behaviors was conducted to compare the relative levels of suicidality in the population. Most items were either binary or categorical. The questionnaire was pilot-tested for its reliability and validity. After minor modifications were made, the questionnaires were then distributed to all the eligible students in the participating schools.

Statistical Analysis

The quantitative data were analyzed using SPSS (version 27.0; IBM Corp). Data cleansing followed the recommendations on treating univariate and multivariate outliers [57]. The results of missing values analysis indicated satisfaction on criteria (Little test) for missing completely at random (χ^2_1=2.3; P=.13), and missing values were replaced through the expectation maximization method. To summarize the distribution of responses on each item, descriptive statistics were presented by frequencies and percentages. Sensitivity analyses using the Pearson chi-square test were performed to examine whether differences in Close Friends use and help-seeking willingness were attributable to respondents' backgrounds or suicide-related experiences. We also used ordinal logistic regression to estimate odds ratios (ORs), adjusted ORs (AORs), and 95% CIs for the associations of the frequency of Close Friends usage with willingness of help seeking and suicidality. Statistical significance was indicated with a P value <.05.

Ethics Approval

Ethical approval was obtained from the local Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster (approval number UW 18-338).

Results

Qualitative Findings

Use of Close Friends and Themes Identified

The usage of Close Friends was mentioned in half of the individual interviews, and 2 out of 6 focus groups addressed the concern regarding public online expression and emotional sharing. Participants who had no experience with private online expression indicated a preference for seeking emotional support from close friends.

The following 3 recurrent themes emerged from the analysis with regard to the usage of Close Friends: (1) reasons for using Close Friends, (2) general concerns about privacy issues, and (3) the importance of seeking help from Close Friends for emotional problems. Each theme is discussed in detail in the next section and the narrative is supported by the illustration of pertinent quotes. Study participants are identified by the interview type and number, gender ("M" and "F" refer to males and females, respectively), age, and suicide-related experiences ("with SI/no SI" indicating whether they had any SI; attempter/nonattempter indicates SA experiences).

Reasons for Using Close Friends: Results From Individual Interviews

Overview

The 3 main reasons for using Close Friends were identified from the individual interviews. These reasons included (1) interaction and help seeking, (2) negative emotions release, and (3) ventilation and self-expression. Relevant quotes are selected and illustrated in the following sections.

Interaction and Help Seeking

Close Friends was commonly used by participants to "interact and seek help from friends". Because of the great heterogeneity of viewers, some participants had difficulty getting support via public posts, while others were hesitant to disturb friends via direct messaging. Therefore, posting on Close Friends turned out to be an ideal option for sharing problems or challenges. It could be an advance notice for friends in real life that a follow-up discussion on the issue was anticipated when they met face-to-face the following day.

> *I never post public stories but always use Close Friends when I post stories on Instagram. There are many people I don't know on this platform. I don't want them to know how I feel. Also, they may not know me, so it is meaningless for them to see the post. Sometimes I can't tell my friends right away after arguing with my dad at night, so I would post [to Close Friends] on Instagram. Or maybe there is nothing serious, and I just want to mention it to them. We can discuss it face to face directly the next day.* [Individual-12, M, 15 years, secondary school student, no SI, nonattempter]

Negative Emotions Release

Close Friends is also a haven for those who are determined to build a favorable public image on social media as it provides a private space for them to "release negative emotions." Indeed, several participants were concerned about how their posts might be treated and whether their posts with emotional expression would demote their prestige, particularly in the eyes of strangers. Thus, using Close Friends made it psychologically safer to publish unfavorable information on SNSs.

> *I prefer to post photos with individualized characteristics to bring a positive feeling to others, and I won't post any negative stuff on social media. However, thanks to Instagram Stories and a new feature called Close Friends, I share more about my daily life [as well as negative emotions]. After all, Instagram is a popular platform with a large number of targeted audiences. I don't want people to think I'm too negative [so I won't post negative things in public].* [Individual-5, F, 18 years, university freshman, with SI, nonattempter]

Ventilation and Self-expression

In addition, as Close Friends posts can only be viewed and commented on by a limited number of followers, some participants believed Close Friends was a suitable outlet to "ventilate" and facilitated their willingness of online expression.

> *I have seen someone put the image of wrist cutting on the Internet, but I would not do the same thing. I would regard [posting online] as one of the ways to ventilate, and I [tended to] say things in a tactful and restrained manner. I use Instagram, but I only have a few followers. Most of them are my close friends. Sometimes I don't think they could understand me, so I will treat the posts as if I am speaking to myself.* [Individual-4, F, 16 years, secondary school student, with SI, attempter]

General Concerns of Privacy Issues: Results From Focus Groups

However, none of the participants from the focus group interviews mentioned their experiences of using Close Friends, although a few did address the privacy concerns when expressing emotions on social media.

> *There was no way to ventilate before, because online platforms were poorly developed then. But now, even if [online platforms are much better developed comparatively], when you have something to share, something you don't like, or you are uncomfortable with, you would choose a group of familiar close friends [instead of everyone online], that is, you will share it in a small circle.* [Group 5: Participant-4, M, 18 years, secondary school student, no SI, nonattempter]

> *IG (Instagram) posts may be viewed by too many people, so I won't post [my status] on it. I will talk to close friends [if I have something to share] by WhatsApp message.* [Group 6: Participant-4, F, 16 years, secondary school student, no SI, nonattempter]

Help Seeking: Importance of Contacting Close Friends for Emotional Problems

Close Friends is a relatively new feature on Instagram; consequently, some participants might not have known about it at the time of the interviews. This could explain why some participants highlighted the importance of contacting and seeking help from close friends but had never actually used Close Friends themselves. It is therefore possible that Close Friends might change their attitudes toward online expression and that these participants might have a greater interest in posting on Instagram after learning about the function.

> *You should find some close friends to chat with, but not with some people who don't know you entirely. A close friend means someone who knows your personality and your ways of doing things. Those who don't know you may only be able to give some poor suggestions.* [Individual-1, F, 16 years, school dropout, with SI, nonattempter]

> *I would talk to my friends about my personal matters in private and I rarely posted the whole story on my Instagram account. I usually shared it in the WhatsApp group since I didn't dare...[directly posting it in public]. I wanted to find someone to listen to me, but I don't like being judged.* [Individual-11, M, 15

years, secondary school student, with SI, nonattempter]

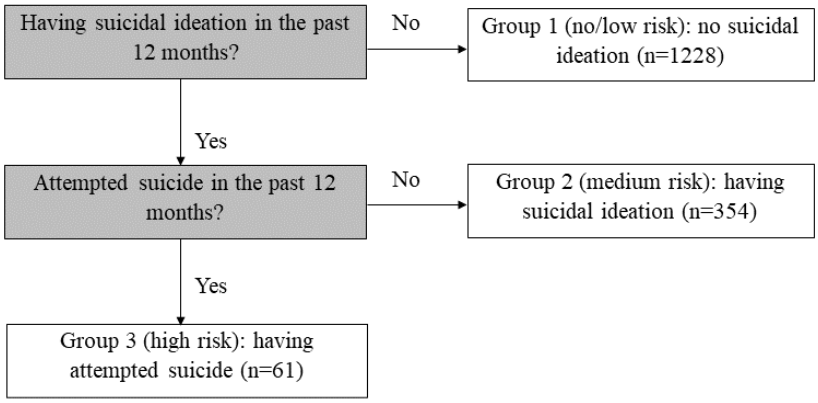
Results of the Questionnaire Survey

Respondents to Questionnaires

Of the 1704 returned questionnaires, 1676 contained valid responses, including 822/1658 from males (49.58%) and 836/1658 from females (50.42%) with a mean age of 16.0 (SD 1.2) years. The number of students in each of the 3 tiers of school bandings was distributed evenly. Bandings are assigned to schools based on students' academic performance in ascending order. Generally speaking, students attending Band 1 schools scored higher on the university entrance examination.

With regard to suicidality, the following 3 groups were identified based on indicated SI and SA experiences over the last 12 months: (1) 1228/1643 (74.74%) respondents reported that they had no SI; (2) 354/1643 (21.55%) respondents reported that they only had SI; and (3) 61/1643 (3.71%) respondents reported that they had attempted suicide. Accordingly, respondents' suicidality was categorized into 3 groups, including "no/low risk" (no SI), "medium risk" (with SI only), and "high risk" (with SA). The procedure of categorization of respondents' suicidality is depicted in Figure 1.

Figure 1. Categorization of the 3 groups for suicidality based on indicated suicide-related experiences.



Frequency of Using Close Friends and Respondents' Help-Seeking Willingness

More than 70% (1163/1646, 70.66%) of the respondents had ever used Close Friends and around 46% (754/1646, 45.81%; 95% CI 0.43-0.48) were frequent users (those who selected "sometimes" and "often"). Around 80% (1010/1266, 79.78%) of the respondents were willing to seek help from peers and friends, yet less than 15% (173/1266, 13.67%) of them indicated an interest in seeking help online for self-distress. The details of respondents' sociodemographic characteristics, help-seeking willingness, and usage of Close Friends are presented in Table 1.

Table 1. Respondents' sociodemographic characteristics, willingness of help seeking, and frequency of using Close Friends (n=1676).

Sociodemographic characteristics	Value, n (%)
Gender (n=1658)[a]	
Male	822 (49.58)
Female	836 (50.42)
School banding	
Band 1	442 (26.37)
Band 2	578 (34.49)
Band 3	656 (39.14)
Willingness of help seeking (n=1266)[b]	
Online	
Yes	173 (13.67)
From peers and friends	
Yes	1010 (79.78)
Frequency of using Close Friends (n=1646)[c]	
Never	483 (29.34)
Seldom	409 (24.85)
Sometimes	492 (29.89)
Often	262 (15.92)

[a]In total, 18 respondents failed to report their genders. These missing values were acceptable considering the size of the entire data set. The valid percentages were thus calculated by 1658 responses.

[b]The valid percentages of willingness to seek help online or from peers and friends were calculated among those who reported willingness to seek help (n=1266).

[c]There were 30 respondents who failed to report the frequency of using Close Friends. These missing values were acceptable considering the size of the entire data set. The valid percentages were thus calculated by 1646 responses.

Differences in the Frequency of Using Close Friends and Help-Seeking Willingness by Respondents' Background Characteristics

Table 2 presents the differences in the frequency of using Close Friends and willingness of help seeking among respondents with varying sociodemographic characteristics and suicide-related experiences. It was found that around 40% (295/802, 36.8%) of male respondents had never used Close Friends, while over half of the female respondents were frequent users (those who answered "sometimes" and "often") of Close Friends (χ^2_3=60.2; *P*<.001). Band 1 school students more often used Close Friends (χ^2_6=12.9; *P*=.04). With regard to help-seeking willingness, females were more active in both online help seeking (χ^2_2=5.8; *P*=.02) and peer and friend–based help seeking (χ^2_1=15.3; *P*<.001). Students from Band 1 schools were less likely to seek help online (χ^2_2=9.3; *P*=.009). Respondents who were willing to seek help from peers and friends were less likely to develop SI (χ^2_2=11.8; *P*=.003), whereas those who were willing to seek help online were more likely to have SI and SA experiences (χ^2_2=9.0; *P*=.01).

Table 2. Willingness of help seeking and the frequency of using Close Friends by respondents' background characteristics (n=1676).[a]

Variables	Gender				School bandings					Suicide-related experiences[b]				
	Male, n/N (%)	Female, n/N (%)	Chi-square (*df*)	*P* value	Band 1, n/N (%)	Band 2, n/N (%)	Band 3, n/N (%)	Chi-square (*df*)	*P* value	Group 1, n/N (%)	Group 2, n/N (%)	Group 3, n/N (%)	Chi-square (*df*)	*P* value
Online help-seeking willingness	67/597 (11.2)	105/660 (15.9)	5.8 (1)	.02	32/345 (9.3)	71/421 (16.9)	70/501 (14.0)	9.3 (2)	.009	117/964 (12.1)	48/247 (19.4)	6/39 (15.4)	9.0 (2)	.01
Peer-oriented help-seeking willingness	450/597 (75.4)	555/659 (84.1)	15.3 (1)	<.001	278/345 (80.6)	340/420 (81.0)	392/501 (78.2)	1.2 (2)	.54	788/963 (81.8)	178/247 (72.1)	30/39 (76.9)	11.8 (2)	.003
Frequency of Close Friends use			60.2 (3)	<.001				12.9 (6)	.04				6.4 (6)	.38
Never	295/802 (36.8)	181/826 (21.9)			110/437 (25.2)	182/574 (31.7)	191/635 (30.1)			349/1205 (29.0)	100/349 (28.7)	20/60 (33.3)		
Seldom	210/802 (26.2)	194/826 (23.5)			97/437 (22.2)	145/574 (25.3)	167/635 (26.3)			315/1205 (26.1)	77/349 (22.1)	12/60 (20.0)		
Sometimes	200/802 (24.9)	288/826 (34.9)			144/437 (33.0)	162/574 (28.2)	186/635 (29.3)			362/1205 (30.0)	110/349 (31.5)	15/60 (25.0)		
Often	97/802 (12.1)	163/826 (19.7)			86/437 (19.7)	85/574 (14.8)	91/635 (14.3)			179/1205 (14.9)	62/349 (17.8)	13/60 (21.7)		

[a]Pearson chi-square tests were used to analyze the data.

[b]The 3 groups were divided for suicidality based on reported suicide-related experiences: group 1 includes those who had no SI, group 2 includes those had SI, and group 3 includes those who had attempted suicide.

Association of the Frequency of Using Close Friends With Help-Seeking Willingness

Table 3 presents the association of the frequency of using Close Friends with online and peer-oriented help-seeking willingness. After adjustment for gender and school banding effect, those who had ever used Close Friends were significantly more (*P*<.001 for "often", *P*<.001 for "sometimes", and *P*=.001 for "seldom") likely to seek help from peers and friends than those who had never used the feature (AORs 1.82-3.02), independent of the frequency with which the feature was used. Respondents who posted in Close Friends most frequently (response of "often") had a higher likelihood (*P*=.03) to seek help online when compared with those who never used the feature (AOR 1.76, 95% CI 1.06-2.93). However, no significant association has been found between the other 2 levels of usage frequency (responses of "sometimes" and "seldom") and online help-seeking willingness (*P*=.44 for "sometimes" and *P*=.16 for "seldom").

Table 3. Summary of ordinal logistic regression analyses for the association of help-seeking willingness with the frequency of using Close Friends (n=1676).[a]

Predictor: Frequency of Close Friends use[b]	Model 1				Model 2[c]			
	Estimate	SE	*P* value	Crude odds ratio (95% CI)	Estimate	SE	*P* value	Adjusted odds ratio (95% CI)
Outcome 1: Willingness of online help seeking [d]								
Frequency: often	0.62	0.25	.02	1.85 (1.13-3.04)	0.57	0.26	.03	1.76 (1.06-2.93)
Frequency: sometimes	0.23	0.24	.33	1.26 (0.79-1.99)	0.19	0.24	.44	1.20 (0.75-1.92)
Frequency: seldom	0.37	0.24	.13	1.44 (0.90-2.31)	0.35	0.24	.16	1.41 (0.88-2.27)
Outcome 2: Willingness of seeking help from peers and friends [e]								
Frequency: often	0.94	0.22	<.001	2.57 (1.65-3.98)	0.86	0.23	<.001	2.36 (1.51-3.70)
Frequency: sometimes	1.13	0.19	<.001	3.10 (2.13-4.51)	1.11	0.20	<.001	3.02 (2.06-4.43)
Frequency: seldom	0.58	0.18	.002	1.78 (1.24-2.55)	0.60	0.19	.001	1.82 (1.27-2.63)

[a]Ordinal logistic regression analyses were used to analyze the data.

[b]For the predictor in both analyses, the frequency of using Close Friends has 4 outcome levels in ascending order: never, seldom, sometimes, and often. The response of "never" was chosen as the reference category.

[c]Model 2 adjusted for gender and school banding.

[d]For outcome 1: willingness of online help seeking, the response of "no" was chosen as the reference category.

[e]For outcome 2: willingness of seeking help from peers and friends, the response of "no" was chosen as the reference category.

Association of Suicidality With the Frequency of Close Friends Use and Willingness of Help Seeking

Table 4 shows the risk factors for suicidality. After adjustment for the gender and school banding effect, willingness to seek help online was associated with an increased risk of suicidality (AOR 1.50, 95% CI 1.04-2.15), while willingness to seek help from peers and friends was associated with a decreased risk of suicidality (AOR 0.55, 95% CI 0.39-0.75). In terms of Close Friends usage frequency, those who "sometimes" used Close Friends had an elevated risk of suicidality compared with those who had never used Close Friends (AOR 1.53, 95% CI 1.01-2.31).

Table 4. Summary of ordinal logistic regression analyses for risk of suicidality (n=1676).[a]

Outcome: risk of suicidality[b]	Model 1				Model 2[c]			
	Estimate	SE	*P* value	Crude odds ratio (95% CI)	Estimate	SE	*P* value	Adjusted odds ratio (95% CI)
Predictor 1: Frequency of Close Friends use (Reference: never)								
Often	0.47	0.21	.03	1.60 (1.06-2.41)	0.40	0.22	.07	1.49 (0.98-2.28)
Sometimes	0.44	0.21	.03	1.56 (1.04-2.34)	0.42	0.21	.047	1.53 (1.01-2.31)
Seldom	0.19	0.19	.34	1.21 (0.82-1.76)	0.23	0.20	.25	1.26 (0.85-1.85)
Predictor 2: Online help-seeking willingness (Reference: no)								
Yes	0.46	0.18	.01	1.59 (1.11-2.27)	0.40	0.19	.03	1.50 (1.04-2.15)
Predictor 3: Peer-oriented help-seeking willingness (Reference: no)								
Yes	–0.58	0.16	<.001	0.56 (0.41-0.77)	–0.61	0.17	<.001	0.55 (0.39-0.75)

[a]Ordinal logistic regression analyses were used to analyze the data.

[b]Risk of suicidality had 3 outcome levels: "no ideation," "having SI," and "having attempted suicide." The response of "no ideation" was chosen as the reference category.

[c]Model 2 adjusted for gender and school banding.

Discussion

Principal Findings

To the best of our knowledge, this is the first study on Close Friends, a newly introduced feature of Instagram. This is also the first rigorous evaluation to explore how the private expression features on SNSs may influence adolescents' willingness of online expression and help seeking with regard to mental distress and suicidality. Our findings demonstrated that a sizable proportion (1163/1646, 70.66%) of adolescents had ever used Close Friends, with around half of the respondents

(754/1646, 45.81%) being frequent users. We identified 3 major motivations for using Close Friends during interviews, including (1) interaction and help seeking, (2) release of negative emotions, and (3) ventilation and self-expression. In terms of help-seeking willingness, youths were largely positive toward seeking help from peers and friends offline, yet negative toward online help seeking from online professionals or online friends with whom they had not yet established a relationship in real life. Most notably, we identified a positive association between the frequency of using Close Friends and the willingness to seek help from peers and friends, as well as a tendency for those who heavily used Close Friends to be predisposed to a higher suicide risk and greater online help-seeking willingness, compared with nonusers. Willingness to seek help online was shown to be positively correlated with suicidality, whereas willingness to seek help from peers and friends was found to be negatively associated with suicidality. The findings of our study, therefore, significantly contribute to the field by highlighting the emerging trend of private online expression and its relationship with suicidality and help-seeking willingness among adolescents. Moreover, it will be effective to raise awareness of suicide prevention in the public so that we can all learn to be the guardian angels of others on social media.

Prevalence of Close Friends Usage and Sociodemographic Variations in Their Use

Our findings show a relatively high prevalence of Close Friends usage among youth. The popularity of Close Friends is as predicted, because personal disclosure has been one of the central aspects of SNS. A higher acceptance rate was found on Instagram among adolescents for distress expression, owing to its advanced privacy settings [58]. The gender differences identified were in line with what was found in the majority of previous studies [59,60]. Interestingly, students from Band 1 schools used Close Friends more often, which contradicted the findings of most prior studies, where social media usage had little or a negative association with academic performance [61,62]. This could be attributable to the study populations (university students vs secondary school students) and intents of using SNSs (entertainment vs private conversation).

Private Online Expression, Willingness of Help Seeking, and Suicidality

Understanding the topic of private online expression and its association with help-seeking willingness and suicidality remains in a nascent state [63]. Our findings suggest that the new feature could substantially facilitate offline help seeking. Notwithstanding, the higher-frequency use of Close Friends may promote online help seeking, whereas the risk of suicidality would increase concurrently. This finding contrasted with previous research indicating that only passive use of SNS (ie, viewing posts), but not active, was linked with a decrease in one's subjective well-being [64,65], and suggested that the frequency of SNS usage could be the determining factor that resulted in the differences. While some studies reported that using SNSs may alleviate loneliness and enhance happiness, others remarked that excessive usage of SNS and online expression can exacerbate the sense of loneliness and impair one's well-being [66]. Similarly, although using Close Friends for private online communication may help strengthen real-world social connections and maintain better social capital, heavy use of the feature may suggest underlying psychological or social malfunction, such as social media addiction, smartphone dependency, lack of confidence in person-to-person interaction, and low self-esteem, all of which indicate poorer mental health [67]. In addition, there is conflicting evidence on the relationship between time spent on SNSs and online help seeking for suicidality [38,40,68], implying that other purposes and motivations of using SNSs may contribute to the variations of both public and private online expression. Considering the basic binary categorization of passive and active use of SNS in most previous studies, future studies may explore the purposes of using SNSs in each individual construct embodied by multiple components.

In other respects, consistent with previous research [69,70], we identified a higher willingness of seeking help from peers and friends and a lower willingness of online help seeking. Besides the lack of knowledge and mental health literacy, based on what interview participants stressed, privacy concerns and the priority of close friends for personal emotions may account for the overall preferences of offline help seeking. Our findings on the association between suicidality and help-seeking willingness are also supported by empirical evidence. For example, previous studies revealed that peer-related loneliness was positively associated with nonsuicidal self-injury (NSSI) engagement [71] and that Chinese adolescents with online help-seeking behaviors had a greater lifetime prevalence of SI [72]. With specific regard to online help seeking, previous research indicated that youngsters with lower life satisfaction and a higher level of stress are more prone to seeking help online [73]. By contrast, although our study provides no evidence of the significant relationship between willingness of online and offline help seeking, a previous study reported that young people who had sought help online for suicide-related issues were less likely to disclose to someone offline [39]. Given that vulnerable youth often report a higher probability of being isolated by peers and alienated from social circles [74], this might explain why they have to turn to people online for help and support. The support from Close Friends could even be more crucial among this group of users.

Implications for Future Studies

Considering the high prevalence of Close Friends among adolescents, this kind of private online expression may shape the behavioral patterns of help seeking. Close Friends provides a secure space for private communication and online expression, which encourages at-risk adolescents to be more authentic in self-disclosure of distress and identity exploration. Increased self-disclosure with intimate friends would, in turn, reciprocally enhance the quality of friendship, facilitate relationship development, and lead to stronger and more stable peer support [75]. Nonetheless, the high level of confidentiality inherent in the Close Friends feature may be challenging for online help services. Albeit previous research shed light on the efficacy of suicide prevention messaging [76] and professional-led online risk screening [72], with the use of Close Friends some negative expressions on SNSs would be circulated within sealed social circles where external access is entirely prohibited. As a

consequence, external service providers would have a little chance to view the messages requesting assistance, and the private use of SNS would make it more controversial for professionals to access and gather data from individuals' social media due to ethical and privacy concerns [77]. In addition, similar to Snapchat and discussion boards [78], Close Friends posts are less content visible and bear more information-sharing affordances. Qualitative findings demonstrate that negative emotional release and ventilation are the 2 primary motivations for youth to initiate expression on Close Friends, and having a private profile (only visible to "friends") on Facebook has been shown to be negatively associated with social capital [79]. Therefore, adolescents who are deeply engaged in posting on Close Friends may be more exposed to peer hostility and contentious comments. Given the correlation between higher-frequency use of Close Friends and online help seeking, researchers and program administrators should take the use of Close Friends into careful consideration and propose solutions to the "blocking" situation.

This study adds evidence to the critical role of peers and friends in youths' help seeking and suicide prevention. A study on friend SNS underscored the value of positive peer evaluation on SNSs for the social adjustment of adolescents [80]. Peer and friend support may reassure at-risk adolescents that they are understood by someone they trust, and even deter them from ongoing or subsequent suicidal behaviors. It is critical to promote or foster some good practices among Close Friends users that encourage them to be attentive to one another's needs, and to seek help from outside if a situation within Close Friends requires immediate attention for the sake of its members' safety. By contrast, although most adolescents intend to help peers in crisis when they read their NSSI posts, some argued that the peer support was not very useful and had little bearing on the decrease in their actual NSSI [81]. One probable explanation is the absence of further guidance in professional consultation, as adolescents often inquire about professionals' suggestions for a peer's condition, but rarely urge the individual to seek formal help directly [82]. Therefore, greater emphasis should be garnered on ways to improve peer training interventions and advocate for peer support services such that everyone can look after each other by providing timely mutual support.

Limitations

This study has a number of limitations. First, the Close Friends feature was not released until midway through the focus groups, which limited how much information we were able to obtain on private expression. Considering suicide is usually a stigmatized topic, the more anonymous data collection approach (eg, online interviews and online surveys) would generate responses with greater validity compared with face-to-face interviews. Second, a validated, multi-item instrument is required in future research that evaluates suicidality and help-seeking willingness. This study included only basic measuring items and demonstrated a simple 3-level structure to indicate youths' suicide risks. Given the mixed relationship between suicidal thoughts, past SAs, and future suicide risk, such a design inferred a certain degree of invalidity. In addition, the use of binary items greatly omitted the details regarding youths' help-seeking willingness and reduced the validity of that information. The result of logistic rather than linear regression could only be considered "preliminary." To expand our knowledge of the nuances of these associations, validated tools that thoroughly assess help-seeking and suicidality are needed. Third, private online expression is a new field of research with no validated questionnaire or scale currently available. Therefore, the items in our survey were derived mainly from qualitative results and the literature and should be revised and tested recurrently in future studies on a similar theme. Additional confounding factors should also be measured, including the number of close friends, relationship issues, parental and school support, and habits of using smartphones and SNSs. Moreover, the number of valid responses to the questionnaire survey fell short of the expected sample size due to an unanticipated amount of missing data and the withdrawal of some schools at the last moment.

Conclusions

The popularity of Close Friends represents the proliferated need for private online expression, reflecting an emerging trend among young people for exchanging suicide-related information. This study demonstrates support for Close Friends usage for self-expression and private conversation among Hong Kong adolescents aged 15-19 years and indicates the relevance and insufficiency of current peer support for suicidal youth. Further studies should be conducted to determine the causal relationship between the frequency and purposes of using Close Friends and willingness to seek help, which would provide more information for the development of suicide prevention initiatives. Researchers and social media platforms should exercise caution when considering the impacts of heavy Close Friends usage and may also collaborate to co-design a risk monitoring system adapted to the private SNS context. Such a system would need to ensure that adolescents' privacy is not jeopardized when communicating online as well as efficiently assist professionals in identifying young people at a high risk of suicide and notify them of any suicide-related information posted online.

Acknowledgments

This study was funded by the Public Policy Research Funding Scheme, Policy Innovation and Co-ordination Office, HKSAR (Project Number 2017.A8.075). The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation of the manuscript. We express our sincere appreciation to Dr Laura Bedford for her kind support with language editing and diligent proofreading of this manuscript.

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. Most popular social networks worldwide as of January 2022, ranked by number of monthly active users (in millions). Statista. 2022. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2022-09-08]
2. Digital 2022: Hong Kong. Datareportal. 2022. URL: https://datareportal.com/reports/digital-2022-hong-kong [accessed 2022-09-12]
3. Distribution of Instagram users worldwide as of April 2022, by age group. Statista. 2022. URL: https://www.statista.com/statistics/325587/instagram-global-age-group/ [accessed 2022-09-12]
4. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital. PLoS One 2018;13(5):e0197712 [FREE Full text] [doi: 10.1371/journal.pone.0197712] [Medline: 29795637]
5. Prescott J, Hanley T, Ujhelyi K. Peer Communication in Online Mental Health Forums for Young People: Directional and Nondirectional Support. JMIR Ment Health 2017 Aug 02;4(3):e29 [FREE Full text] [doi: 10.2196/mental.6921] [Medline: 28768607]
6. Dyson MP, Hartling L, Shulhan J, Chisholm A, Milne A, Sundar P, et al. A Systematic Review of Social Media Use to Discuss and View Deliberate Self-Harm Acts. PLoS One 2016;11(5):e0155813 [FREE Full text] [doi: 10.1371/journal.pone.0155813] [Medline: 27191728]
7. Lewis SP, Seko Y. A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. J Clin Psychol 2016 Mar;72(3):249-262. [doi: 10.1002/jclp.22242] [Medline: 26613372]
8. Nasier B, Gibson K, Trnka S. "PM me" or "LOL": Young peoples' observations of supportive and unsympathetic responses to distress on social media. Computers in Human Behavior 2021 Nov;124:106933 [FREE Full text] [doi: 10.1016/j.chb.2021.106933]
9. Forest AL, Wood JV. When social networking is not working: individuals with low self-esteem recognize but do not reap the benefits of self-disclosure on Facebook. Psychol Sci 2012 Mar;23(3):295-302. [doi: 10.1177/0956797611429709] [Medline: 22318997]
10. Bazarova N, Taft J, Choi Y, Cosley D. Managing Impressions and Relationships on Facebook. Journal of Language and Social Psychology 2012 Aug 30;32(2):121-141 [FREE Full text] [doi: 10.1177/0261927x12456384]
11. Valkenburg P, Peter J. Social Consequences of the Internet for Adolescents. Curr Dir Psychol Sci 2009 Feb 01;18(1):1-5 [FREE Full text] [doi: 10.1111/j.1467-8721.2009.01595.x]
12. Frison E, Eggermont S. Toward an Integrated and Differential Approach to the Relationships Between Loneliness, Different Types of Facebook Use, and Adolescents' Depressed Mood. Communication Research 2015 Dec 03;47(5):701-728 [FREE Full text] [doi: 10.1177/0093650215617506]
13. Boyd D, Ellison N. Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication 2007;13(1):210-230 [FREE Full text] [doi: 10.1111/j.1083-6101.2007.00393.x]
14. Chan G. Intimacy, friendship, and forms of online communication among hidden youth in Hong Kong. Computers in Human Behavior 2020 Oct;111:106407 [FREE Full text] [doi: 10.1016/j.chb.2020.106407]
15. Agosto D, Abbas J. "Don't be dumb—that's the rule I try to live by": A closer look at older teens' online privacy and safety attitudes. New Media & Society 2016 Jul 09;19(3):347-365 [FREE Full text] [doi: 10.1177/1461444815606121]
16. Marwick A, Boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society 2010 Jul 07;13(1):114-133 [FREE Full text] [doi: 10.1177/1461444810365313]
17. Budenz A, Klassen A, Purtle J, Yom-Tov E, Yudell M, Massey P. "If I was to post something, it would be too vulnerable:" University students and mental health disclosures on instagram. J Am Coll Health 2022;70(2):615-624. [doi: 10.1080/07448481.2020.1759608] [Medline: 32407177]
18. Sofia D, Schinria I, Elizabeth R, Niloufar S. Finsta: Creating "Fake" Spaces for Authentic Performance. In: CHI '19: Proceedings 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. 1-6. [doi: 10.1145/3290607.3313033]
19. Kang J, Wei L. Let me be at my funniest: Instagram users' motivations for using Finsta (a.k.a., fake Instagram). The Social Science Journal 2020 Jan 27;57(1):58-71 [FREE Full text] [doi: 10.1016/j.soscij.2018.12.005]
20. Peter J, Valkenburg P. Research Note: Individual Differences in Perceptions of Internet Communication. European Journal of Communication 2016 Jul 26;21(2):213-226 [FREE Full text] [doi: 10.1177/0267323105064046]
21. Meta. Instagram Stories. Instagram. 2021. URL: https://business.instagram.com/instagram-stories [accessed 2022-09-14]

22. Meta. How do I share a story with my close friends list on Instagram? Instagram. 2021. URL: https://help.instagram.com/2183694401643300 [accessed 2022-09-14]
23. Sihombing L, Aninda M. Phenomenology Of Using Instagram Close Friend Features For Self Disclosure Improvement. Professional: Jurnal Komunikasi dan Administrasi Publik 2022 Jun;9(1):29-34 [FREE Full text]
24. World Health Organization (WHO). Suicide. WHO. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2022-09-14]
25. Yip PSF, Zheng Y, Wong C. Demographic and epidemiological decomposition analysis of global changes in suicide rates and numbers over the period 1990-2019. Inj Prev 2022 Apr;28(2):117-124. [doi: 10.1136/injuryprev-2021-044263] [Medline: 34400542]
26. Pirkis J, Amadeo S, Beautrais A, Phillips M, Yip PSF. Suicide Prevention in the Western Pacific Region. Crisis 2020 Mar;41(Suppl 1):S80-S98. [doi: 10.1027/0227-5910/a000670] [Medline: 32208763]
27. Committee on Prevention of Student Suicides. Final Report November 2016 by Committee on Prevention of Student Suicides. The Government of Hong Kong Special Administrative Region. 2016. URL: https://mentalhealth.edb.gov.hk/uploads/mh/content/document/The_Committee_on_Prevention_of_Student_Suicide_Final_Report_Nov_2016.pdf [accessed 2022-09-14]
28. Yang C, Yip PSF. Changes in the epidemiological profile of suicide in Hong Kong: a 40-year retrospective decomposition analysis. China Popul Dev Stud 2021;5(2):153-173 [FREE Full text] [doi: 10.1007/s42379-021-00087-5] [Medline: 34249402]
29. World Health Organization (WHO). Preventing suicide: A global imperative. WHO. Geneva, Switzerland: WHO; 2014. URL: https://www.who.int/publications/i/item/9789241564779 [accessed 2022-09-14]
30. HKJC Centre for Suicide Research and Prevention. Suicide rates by age groups in Hong Kong, 2011-2021. HKJC Centre for Suicide Research and Prevention. 2022. URL: https://hkjccsrp.hku.hk/statistics/ [accessed 2022-09-14]
31. Elmquist D, McLaughlin C. Social Media Use Among Adolescents Coping with Mental Health. Contemp School Psychol 2017 Nov 16;22(4):503-511 [FREE Full text] [doi: 10.1007/s40688-017-0167-5]
32. Lee E, Lee J, Moon JH, Sung Y. Pictures Speak Louder than Words: Motivations for Using Instagram. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):552-556. [doi: 10.1089/cyber.2015.0157] [Medline: 26348817]
33. Sheldon P, Bryant K. Instagram: Motives for its use and relationship to narcissism and contextual age. Computers in Human Behavior 2016 May;58:89-97 [FREE Full text] [doi: 10.1016/j.chb.2015.12.059]
34. Chen SS, Lam TP, Lam KF, Lo TL, Chao DVK, Mak KY, et al. Motivations for Online Expression, Willingness of Online Help-Seeking, and the Risk of Suicide Among Hong Kong Youths: A Mixed-Methods Study. Cyberpsychol Behav Soc Netw 2022 Jun;25(6):384-391. [doi: 10.1089/cyber.2022.0007] [Medline: 35594242]
35. Robinson J, Bailey E, Witt K, Stefanac N, Milner A, Currier D, et al. What Works in Youth Suicide Prevention? A Systematic Review and Meta-Analysis. EClinicalMedicine 2018;4-5:52-91 [FREE Full text] [doi: 10.1016/j.eclinm.2018.10.004] [Medline: 31193651]
36. Schlichthorst M, Ozols I, Reifels L, Morgan A. Lived experience peer support programs for suicide prevention: a systematic scoping review. Int J Ment Health Syst 2020;14:65 [FREE Full text] [doi: 10.1186/s13033-020-00396-1] [Medline: 32817757]
37. Kitagawa Y, Shimodera S, Togo F, Okazaki Y, Nishida A, Sasaki T. Suicidal feelings interfere with help-seeking in bullied adolescents [corrected]. PLoS One 2014;9(9):e106031 [FREE Full text] [doi: 10.1371/journal.pone.0106031] [Medline: 25188324]
38. Harris KM, McLean JP, Sheffield J. Examining suicide-risk individuals who go online for suicide-related purposes. Arch Suicide Res 2009;13(3):264-276. [doi: 10.1080/13811110903044419] [Medline: 19591000]
39. Bell J, Mok K, Gardiner E, Pirkis J. Suicide-Related Internet Use Among Suicidal Young People in the UK: Characteristics of Users, Effects of Use, and Barriers to Offline Help-Seeking. Arch Suicide Res 2018;22(2):263-277. [doi: 10.1080/13811118.2017.1334609] [Medline: 28590842]
40. Harris KM, McLean JP, Sheffield J. Suicidal and online: how do online behaviors inform us of this high-risk population? Death Stud 2014;38(6-10):387-394. [doi: 10.1080/07481187.2013.768313] [Medline: 24666145]
41. Wright KB, Rosenberg J, Egbert N, Ploeger NA, Bernard DR, King S. Communication competence, social support, and depression among college students: a model of facebook and face-to-face support network influence. J Health Commun 2013;18(1):41-57. [doi: 10.1080/10810730.2012.688250] [Medline: 23030518]
42. Wright K. Emotional Support and Perceived Stress Among College Students Using Facebook.com: An Exploration of the Relationship Between Source Perceptions and Emotional Support. Communication Research Reports 2012 Jul 20;29(3):175-184 [FREE Full text] [doi: 10.1080/08824096.2012.695957]
43. Wong K, Chan CS, Chan M, Wong C, Cheng Q, Xiong C, et al. Who seeks help online? Comparing online and offline help-seeking preferences amongst youths with suicidal ideation. J Affect Disord 2021 Sep 01;292:21-29. [doi: 10.1016/j.jad.2021.05.056] [Medline: 34087633]
44. Lim S, Chan Y, Vadrevu S, Basnyat I. Managing peer relationships online – Investigating the use of Facebook by juvenile delinquents and youths-at-risk. Computers in Human Behavior 2013 Jan;29(1):8-15 [FREE Full text] [doi: 10.1016/j.chb.2012.04.025]

45. Cui S, Cheng Y, Xu Z, Chen D, Wang Y. Peer relationships and suicide ideation and attempts among Chinese adolescents. Child Care Health Dev 2011 Sep;37(5):692-702. [doi: 10.1111/j.1365-2214.2010.01181.x] [Medline: 21198776]
46. D'Avanzo B, Barbato A, Erzegovesi S, Lampertico L, Rapisarda F, Valsecchi L. Formal and informal help-seeking for mental health problems. A survey of preferences of italian students. Clin Pract Epidemiol Ment Health 2012;8:47-51 [FREE Full text] [doi: 10.2174/1745017901208010047] [Medline: 22715343]
47. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015 Mar 07;15:40 [FREE Full text] [doi: 10.1186/s12888-015-0429-6] [Medline: 25886609]
48. Leung L. Stressful life events, motives for Internet use, and social support among digital kids. Cyberpsychol Behav 2007 Apr;10(2):204-214. [doi: 10.1089/cpb.2006.9967] [Medline: 17474837]
49. Yip P, Chan WL, Cheng Q, Chow S, Hsu SM, Law YW, et al. A 24-hour online youth emotional support: Opportunities and challenges. Lancet Reg Health West Pac 2020 Nov;4:100047 [FREE Full text] [doi: 10.1016/j.lanwpc.2020.100047] [Medline: 34327390]
50. Kingsbury M, Reme B, Skogen J, Sivertsen B, Øverland S, Cantor N, et al. Differential associations between types of social media use and university students' non-suicidal self-injury and suicidal behavior. Computers in Human Behavior 2021 Feb;115:106614 [FREE Full text] [doi: 10.1016/j.chb.2020.106614]
51. Vannucci A, McCauley Ohannessian C. Social Media Use Subgroups Differentially Predict Psychosocial Well-Being During Early Adolescence. J Youth Adolesc 2019 Aug;48(8):1469-1493. [doi: 10.1007/s10964-019-01060-9] [Medline: 31256313]
52. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern‐based qualitative analytic approaches. Couns Psychother Res 2020 Oct 18;21(1):37-47 [FREE Full text] [doi: 10.1002/capr.12360]
53. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology 2020 Aug 12;18(3):328-352 [FREE Full text] [doi: 10.1080/14780887.2020.1769238]
54. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597 [FREE Full text] [doi: 10.1080/2159676X.2019.1628806]
55. Student Enrolment Statistics, 2019/20 (Kindergartens, Primary and Secondary Schools). Hong Kong Education Bureau, HKSAR. 2019. URL: https://www.edb.gov.hk/attachment/en/about-edb/publications-stat/figures/Enrol_2019.pdf [accessed 2022-09-14]
56. Siu AMH. Self-Harm and Suicide Among Children and Adolescents in Hong Kong: A Review of Prevalence, Risk Factors, and Prevention Strategies. J Adolesc Health 2019 Jun;64(6S):S59-S64. [doi: 10.1016/j.jadohealth.2018.10.004] [Medline: 31122551]
57. Tabachnick BG, Fidell LS, Ullman JB. Using Multivariate Statistics (5th edition). Boston, MA: Pearson Education; 2007.
58. Kwon S, Park E, Kim K. What drives successful social networking services? A comparative analysis of user acceptance of Facebook and Twitter. The Social Science Journal 2019 Dec 09;51(4):534-544 [FREE Full text] [doi: 10.1016/j.soscij.2014.04.005]
59. Sun B, Mao H, Yin C. Male and Female Users' Differences in Online Technology Community Based on Text Mining. Front Psychol 2020;11:806 [FREE Full text] [doi: 10.3389/fpsyg.2020.00806] [Medline: 32528342]
60. Balt E, Mérelle S, van Bergen D, Gilissen R, van der Post P, Looijmans M, et al. Gender differences in suicide-related communication of young suicide victims. PLoS One 2021;16(5):e0252028 [FREE Full text] [doi: 10.1371/journal.pone.0252028] [Medline: 34019584]
61. AlFaris E, Irfan F, Ponnamperuma G, Jamal A, Van der Vleuten C, Al Maflehi N, et al. The pattern of social media use and its association with academic performance among medical students. Med Teach 2018 Sep;40(sup1):S77-S82. [doi: 10.1080/0142159X.2018.1465536] [Medline: 29732945]
62. Giunchiglia F, Zeni M, Gobbi E, Bignotti E, Bison I. Mobile social media usage and academic performance. Computers in Human Behavior 2018 May;82:177-185 [FREE Full text] [doi: 10.1016/j.chb.2017.12.041]
63. Nesi J, Burke TA, Bettis AH, Kudinova AY, Thompson EC, MacPherson HA, et al. Social media use and self-injurious thoughts and behaviors: A systematic review and meta-analysis. Clin Psychol Rev 2021 Jul;87:102038 [FREE Full text] [doi: 10.1016/j.cpr.2021.102038] [Medline: 34034038]
64. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and Passive Social Media Use and Symptoms of Anxiety and Depressed Mood Among Icelandic Adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
65. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. J Exp Psychol Gen 2015 Apr;144(2):480-488. [doi: 10.1037/xge0000057] [Medline: 25706656]
66. Phu B, Gow A. Facebook use and its association with subjective happiness and loneliness. Computers in Human Behavior 2019 Mar;92:151-159 [FREE Full text] [doi: 10.1016/j.chb.2018.11.020]
67. Ye S, Ho K, Zerbe A. The effects of social media usage on loneliness and well-being: analysing friendship connections of Facebook, Twitter and Instagram. IDD 2021 Feb 22;49(2):136-150 [FREE Full text] [doi: 10.1108/idd-08-2020-0091]

68. Harris KM, Starcevic V, Ma J, Zhang W, Aboujaoude E. Suicidality, psychopathology, and the internet: Online time vs. online behaviors. Psychiatry Res 2017 Sep;255:341-346. [doi: 10.1016/j.psychres.2017.06.012] [Medline: 28601719]
69. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: A systematic narrative review. Children and Youth Services Review 2014 Jun;41:27-36 [FREE Full text] [doi: 10.1016/j.childyouth.2014.03.001]
70. Mar M, Neilson E, Torchalla I, Werker G, Laing A, Krausz M. Exploring e-Mental Health Preferences of Generation Y. Journal of Technology in Human Services 2014 Nov 26;32(4):312-327 [FREE Full text] [doi: 10.1080/15228835.2014.943457]
71. Gandhi A, Luyckx K, Goossens L, Maitra S, Claes L. Association between Non-Suicidal Self-Injury, Parents and Peers Related Loneliness, and Attitude Towards Aloneness in Flemish Adolescents: An Empirical Note. Psychol Belg 2018 Feb 20;58(1):3-12 [FREE Full text] [doi: 10.5334/pb.385] [Medline: 30479803]
72. Chan M, Li TMH, Law YW, Wong PWC, Chau M, Cheng C, et al. Engagement of vulnerable youths using internet platforms. PLoS One 2017;12(12):e0189023 [FREE Full text] [doi: 10.1371/journal.pone.0189023] [Medline: 29261687]
73. Utz S, Breuer J. The Relationship Between Use of Social Network Sites, Online Social Support, and Well-Being: Results From a Six-Wave Longitudinal Study. J Media Psychol 2017;29(3):115-125 [FREE Full text] [doi: 10.1027/1864-1105/a000222] [Medline: 29147141]
74. Aguirre Velasco A, Cruz ISS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
75. Mesch G. Social context and communication channels choice among adolescents. Computers in Human Behavior 2009 Jan;25(1):244-251 [FREE Full text] [doi: 10.1016/j.chb.2008.09.007]
76. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing Social Media-Based Suicide Prevention Messages in Partnership With Young People: Exploratory Study. JMIR Ment Health 2017 Oct 04;4(4):e40 [FREE Full text] [doi: 10.2196/mental.7847] [Medline: 28978499]
77. Pourmand A, Roberson J, Caggiula A, Monsalve N, Rahimi M, Torres-Llenza V. Social Media and Suicide: A Review of Technology-Based Epidemiology and Risk Assessment. Telemed J E Health 2019 Oct;25(10):880-888. [doi: 10.1089/tmj.2018.0203] [Medline: 30362903]
78. Renninger B. “Where I can be myself … where I can speak my mind” : Networked counterpublics in a polymedia environment. New Media & Society 2014 Apr 09;17(9):1513-1529 [FREE Full text] [doi: 10.1177/1461444814530095]
79. Stutzman F, Gross R, Acquisti A. Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. JPC 2013 Mar 01;4(2):620 [FREE Full text] [doi: 10.29012/jpc.v4i2.620]
80. Livingstone S. Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. New Media & Society 2008 Jun 01;10(3):393-411 [FREE Full text] [doi: 10.1177/1461444808089415]
81. Brown R, Fischer T, Goldwich D, Plener P. "I just finally wanted to belong somewhere"-Qualitative Analysis of Experiences With Posting Pictures of Self-Injury on Instagram. Front Psychiatry 2020;11:274 [FREE Full text] [doi: 10.3389/fpsyt.2020.00274] [Medline: 32372983]
82. Curtis C. Youth perceptions of suicide and help-seeking: ‘They'd think I was weak or “mental”’. Journal of Youth Studies 2010 Jul 28;13(6):699-715 [FREE Full text] [doi: 10.1080/13676261003801747]

Abbreviations

AOR: adjusted odds ratio
NSSI: nonsuicidal self-injury
OR: odds ratio
SA: suicide attempt
SI: suicidal ideation
SNS: social networking site

Edited by R Kukafka; submitted 03.03.22; peer-reviewed by K Harris, J Breuer; comments to author 28.05.22; revised version received 20.07.22; accepted 06.09.22; published 12.10.22.

Please cite as:
Chen SS, Lam TP, Lam KF, Lo TL, Chao DVK, Mak KY, Lam EWW, Tang WS, Chan HY, Yip PSF
The Use of Close Friends on Instagram, Help-Seeking Willingness, and Suicidality Among Hong Kong Youth: Exploratory Sequential Mixed Methods Study
J Med Internet Res 2022;24(10):e37695
URL: https://www.jmir.org/2022/10/e37695
doi:10.2196/37695
PMID:36223182

Original Paper

Assessment of the Popularity and Perceived Effectiveness of Smartphone Tools That Track and Limit Smartphone Use: Survey Study and Machine Learning Analysis

Elias Aboujaoude[1], MD; Germano Vera Cruz[2], PhD; Lucien Rochat[3], PhD; Robert Courtois[4], MD, PhD; Farah Ben Brahim[4], MSc; Riaz Khan[5], Prof Dr; Yasser Khazaal[6], MD

[1]Stanford University School of Medicine, Department of Psychiatry and Behavioral Sciences, Stanford, CA, United States

[2]Department of Psychology, University of Picardie Jules Verne, Amiens, France

[3]Addiction Division, Department of Psychiatry, University Hospitals of Geneva, Geneva, Switzerland

[4]Department of Psychology, Tours University, Tours, France

[5]Addiction Psychiatry, Foederatio Medicorum Helveticorum, Geneva, Switzerland

[6]Lausanne University, Lausanne, Switzerland

Corresponding Author:
Elias Aboujaoude, MD
Stanford University School of Medicine
Department of Psychiatry and Behavioral Sciences
401 Quarry Rd
Stanford, CA, 94305
United States
Phone: 1 650 498 9111
Email: eaboujaoude@stanford.edu

Abstract

Background: Problematic smartphone use, like problematic internet use, is a condition for which treatment is being sought on the web. In the absence of established treatments, smartphone-provided tools that monitor or control smartphone use have become increasingly popular, and their dissemination has largely occurred without oversight from the mental health field.

Objective: We aimed to assess the popularity and perceived effectiveness of smartphone tools that track and limit smartphone use. We also aimed to explore how a set of variables related to mental health, smartphone use, and smartphone *addiction* may influence the use of these tools.

Methods: First, we conducted a web-based survey in a representative sample of 1989 US-based adults using the crowdsourcing platform Prolific. Second, we used machine learning and other statistical tools to identify latent user classes; the association between latent class membership and demographic variables; and any predictors of latent class membership from covariates such as daily average smartphone use, social problems from smartphone use, smartphone *addiction*, and other psychiatric conditions.

Results: Smartphone tools that monitor and control smartphone use were popular among participants, including parents targeting their children; for example, over two-thirds of the participants used sleep-related tools. Among those who tried a tool, the highest rate of perceived effectiveness was 33.1% (58/175). Participants who experienced problematic smartphone use were more likely to be younger and more likely to be female. Finally, 3 latent user classes were uncovered: nonusers, effective users, and ineffective users. Android operating system users were more likely to be nonusers, whereas younger adults and females were more likely to be effective users. The presence of psychiatric symptoms did not discourage smartphone tool use.

Conclusions: If proven effective, tools that monitor and control smartphone use are likely to be broadly embraced. Our results portend well for the acceptability of mobile interventions in the treatment of smartphone-related psychopathologies and, potentially, non–smartphone-related psychopathologies. Better tools, targeted marketing, and inclusive design, as well as formal efficacy trials, are required to realize their potential.

(J Med Internet Res 2022;24(10):e38963) doi:10.2196/38963

KEYWORDS

smartphone addiction; internet addiction; internet gaming disorder; smartphone tools; telepsychiatry; machine learning; telemedicine; social media; digital mental health interventions; mobile phone

Introduction

Background

The recognition of psychological downsides to internet-related technologies is >2 decades old. A large body of epidemiological, phenomenological, and biological research has accumulated during that period, leading to the inclusion of *Gaming Disorder* in the International Classification of Diseases, 11th revision [1], and of *Internet Gaming Disorder* in the appendix to the Diagnostic and Statistical Manual of Mental Disorders (DSM), 5th Edition [2]. Although the field may have become better at identifying internet-related psychopathology and, in some cases, attaching an accepted diagnostic label to it, agreed upon treatment guidelines remain elusive.

Conventional Treatments

Psychopharmacological interventions have been inspired by conditions to which problematic internet use has been compared, including obsessive-compulsive disorder, substance use disorders, behavioral addictions, and attention deficit hyperactivity disorder [3]. However, the relatively limited exploration of serotonin reuptake inhibitors [4], mu receptor antagonists [5], and stimulants [6] has not yielded solid evidence to support their broad use. Psychotherapeutic interventions—individual, group, and residential—have received more research attention [7] and possess a larger evidence base, especially in favor of cognitive behavioral therapy. However, methodological differences, sample nonrepresentativeness, and other research study limitations preclude strong conclusions and recommendations for wider adoption.

Using Technology Against Itself

In this relative treatment vacuum, and in parallel with growing social and cultural recognition of the risks to personal well-being of runaway smartphone reliance, a new help modality has emerged and rapidly asserted itself among technology users and developers, with little direct contribution to its growth and design from clinicians and mental health experts. Described as "using technology against itself" [8], it involves 2 basic offerings: functionalities built into the smartphone that can be activated at will to monitor and limit use; and apps (dubbed "apps to wean us off apps") that can be downloaded from a third party and used for the same purpose. Like the *old* psychopharmacological and psychotherapeutic interventions explored for problematic internet use, the goal is to track and curtail excessive or problematic smartphone use through tools that, if proven successful, might possess some unique advantages over traditional interventions, including scalability, cost-effectiveness, diminished stigma, convenience, and lack of side effects [8]. Aligned with mobile therapy, these interventions may also benefit from the big acceptability gains that the telepsychiatry field has enjoyed among patients and providers during the COVID-19 pandemic, further propelling their growth in the years to come.

This study assessed the use of built-in smartphone tools meant to monitor and target problematic or excessive smartphone use among US-based users. Examples of such tools include screen time (tracks and quantifies use), grayscale (makes apps and alerts less noticeable), disabling notifications (reduces distractions), audio messaging (limits typing and reduces confusion), night shift (reduces exposure to sleep-disrupting blue light), tools that move apps from the home screen to reduce distractibility, and tools that delete apps. Examining the popularity of such tools, users' experiences with them, and any associations with demographic, mental health, and psychosocial factors can shed light on the promise and limitations of this growing field and help start a much-needed assessment by the mental health community of its place within the broader treatment landscape.

To do so, this study aimed to assess the prevalence of problematic smartphone use in a representative sample of US-based users, as well as the use of tools to monitor and control smartphone use (TMCSU). The study also aimed to explore how a set of variables related to mental health, smartphone use, and smartphone *addiction* may influence the use of TMCSU.

Toward that goal, we tried to answer 6 research questions (RQs), presented in Textbox 1.

Textbox 1. Research questions.

Research questions regarding a set of variables affecting use of tools to monitor and control smartphone use (TMCSU):

1. Research question (RQ) 1: In a representative sample of users in the United States, what is the prevalence of problematic smartphone use, as measured by variables such as daily average use, daily average use for nonessential activities, social problems because of smartphone use, and smartphone *addiction*?
2. RQ2: In a representative sample of users in the United States, how common is the use of TMCSU and what is the perceived effectiveness of these tools? The tools are designed to, for instance, measure daily smartphone use duration, manage notifications, block problematic apps, improve sleep time and quality, and monitor and control smartphone use in underage children.
3. RQ3: In a representative sample of users in the United States, can we identify the underlying latent classes of smartphone users using TMCSU?
4. RQ4: If the underlying latent classes of smartphone users are identified, what are the associations between the participants' sociodemographic characteristics (age and sex) and the uncovered latent classes?
5. RQ5: Among variables such as participants' smartphone operating system, daily average smartphone use, social problems owing to smartphone use, smartphone *addiction*, and diagnosed mental health disorders, what are the most important features (covariates) predicting smartphone users' latent class membership?
6. RQ6: What are the associations between the covariates (smartphone operating system, daily average smartphone use, social problems owing to smartphone use, smartphone *addiction*, mental health disorders, etc) and the latent classes associated with the use of TMCSU?

As this was designed as an exploratory study, no hypotheses were associated with these RQs.

To our knowledge, and despite the popularity of smartphone tools intended to curb smartphone use, this is the first psychological assessment of this digital mental health intervention. As such, our findings can help guide a field that is growing rapidly but mostly outside of meaningful scrutiny by the mental health scientific community.

Methods

Participants

A representative sample of the adult population in the United States was recruited. Overall, 1989 individuals participated in the survey and answered a web-based questionnaire. To be included, participants had to be adults (≥18 years), belong to a Prolific representative sample of the adult population in the United States, and provide digital informed consent for study participation. The only exclusion criterion was reporting no smartphone use.

The 2 sociodemographic characteristics collected were age and sex. Age varied between 19 and 76 years (mean 45, SD 16 years) and was distributed as follows: 19 to 35 years (691/1989, 34.74%); 36 to 60 years (802/1989, 40.32%); and 61 to 76 years (496/1989, 24.93%). The sex distribution was as following: male (965/1989, 48.51%); female (1006/1989, 50.57%); and nonbinary (18/1989, 0.9%).

Recruitment and Sampling

The recruitment of study participants was conducted anonymously using the web-based crowdsourcing platform Prolific [9]. Prolific has been described as possessing some advantages over other similar platforms, including that it is exclusively dedicated to research studies, and its participants are more ethnically and geographically diverse and naive to experimental research tasks [10]. As such, it can allow the recruitment of a US-based sample of adults aged ≥18 years with sex, age, and ethnicity characteristics that reflect the US Census Bureau data. According to the Prolific platform, only people who have an account on Prolific are notified of studies that they are eligible for based on the demographic information they provide. When the study was posted on Prolific, an invitation was sent by Prolific to a random subset of all eligible individuals. To be eligible for a US representative sample, Prolific participants must be residents of the United States and be fluent in English. The sample in this study is approximately the maximum deliverable representative sample size by Prolific. Of the whole sample, 2% were excluded from the analyses because they did not use a smartphone. Participants received approximately US $1.3 for completing the questionnaire. The recruitment was launched and completed in March 2021.

Data Collection Material

Data were collected using a web-based questionnaire. The questionnaire included 80 questions divided into 5 parts. The first part included 48 questions assessing sociodemographic characteristics (eg, age and sex); a screening question about smartphone use; items on smartphone operating system (OS); items on smartphone use behavior during the previous 12 months (eg, daily use duration, use behavior, daily average use duration on nonessential activities, and social problems owing to smartphone use); items on the use of smartphone TMCSU behavior during the previous 12 months (whether participants used them, and, if yes, how often [rarely, sometimes, or frequently] and with what perceived effectiveness [very effective, somewhat effective, neither effective nor ineffective, somewhat ineffective, or very ineffective]); and items on lifetime mental health and alcohol use disorder diagnoses (yes or no).

The second and third parts of the questionnaire assessed mental health and smartphone addiction, respectively. The latter was assessed using the smartphone application-based addiction scale [11], which comprises 6 items and a 6-point response scale ranging from *strongly disagree* to *strongly agree*. Mental health–related symptoms were assessed using the DSM, 5th Edition self-rated level 1 cross-cutting symptom measure [12] (24 items and a 5-point response scale ranging from *not at all* to *nearly every day*), which assesses symptoms over the 2 weeks preceding survey completion across 14 domains: depression, anxiety, anger, mania, personality functioning, sleep problems,

somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviors, and dissociation and substance use. In this study, these dimensions were labeled *DSM Depression*, *DSM Anxiety*, and so on.

The fourth part included a single item on how smartphone use may have changed during the COVID-19 pandemic (7-point response scale ranging from *decrease* to *increase*). Finally, the fifth part of the questionnaire included a single question about daily smartphone screen time. The raw data and collection materials were accessible through a coauthor's research database [13].

Data Analysis

To answer RQ1, we built a cross tab with the problematic smartphone use variables against sociodemographic characteristic variables, and we conducted a Pearson chi-square analysis on the resulting contingency table.

To answer RQ2, descriptive statistical analyses were conducted on the data regarding participants' use of TMCSU.

To answer RQ3, we conducted latent class analysis (LCA) using the *poLCA* R package. LCA is a way to uncover hidden groupings within data. From the participants' item responses, LCA algorithms divided participants into subgroups based on unobservable constructs (latent variables). The resulting subgroups were called latent classes. This technique is analogous to factor analysis where the model determines the latent variables from the manifest (measured) variables. Unlike cluster analysis techniques, which are based on mathematical distances (eg, Euclidean distances) or mathematical density, LCA is based on participants' probability of giving different designed modalities of response (items-response probability) and the probability of membership in the modeled latent classes. Therefore, LCA is considered more advantageous than cluster analysis for model selection and interpretation [14,15]. In total, we ran 9 LCA models, with the first model nclass=2, second model nclass=3, and so on. In each model, the other poLCA function parameter settings were as follows: nrep=10, na.rm=F, graphs=T, and maxiter=100,000. After ensuring that the 9 models built were well identified (through maximum likelihood estimation), we proceeded with the comparison and model selection. On the basis of the Akaike information criterion, the Bayesian information criterion, entropy metrics (respectively 30224.17, 30858.43, and 0.83), and interpretability, we selected the model with 3 latent classes.

To answer RQ4, we computed a cross tab with age against sex against latent class, and then calculated Pearson chi-square of independence.

To answer RQ5, we built a machine learning model using the random forest (RF) classification algorithm [16]. The RF method uses a random subset of predictors and participants and, through recursive partitioning, tests the strength of each available predictor variable individually. This involves building a decision tree from the strongest available predictors and testing the tree's overall predictive power on the *out-of-bag* sample (a subset of data that were not used to build the tree). The RF algorithm performs this repeatedly, separately bootstrapping thousands of decision trees and then averaging them out. RF classification models reveal, among other outputs, the importance of each predictor variable (predictors which made the largest contributions to the model) based on a measure called mean decrease accuracy (MDA). The MDA plot expresses how much accuracy the model loses by excluding each variable. The more the accuracy suffers, the more important the variable is for successful classification. Thus, the variables can be presented in ascending or descending order of importance. RF is nonparametric and, in essence, is able to capture nonlinear relationships [16]. To select the best model, we constructed 4 classification models with different fitting parameters. Each model was built using the *randomForest* package for R. In machine learning, the original data set is split into at least 2 sets: one to train the model (train-set; usually 70%-80% of the sample), the other to estimate the performance of the model when used to make predictions (test-set; 20%-30% of the sample). In this study, the data set was split as follows: train-set=70% and test-set=30% of the sample. The selected model had the following tuning parameters: *ntree*=500, which means that each RF model was built from 500 classification trees; and *mtry*=8, which means that the number of predictors available for splitting at each tree node was set to 8. The performance metrics of the selected model on the test-set data were as follows: accuracy score=0.73 (95% CI 0.66-0.78); no information rate=69; *P* (accuracy>no information rate)=0.038; κ=0.74.

The choice of using machine learning algorithms instead of traditional methods stems from the fact that these algorithms have hyperparameters that can be used to build different models with improved prediction capabilities to test the models' respective performance using a subset of the main data set (named test-set) and to choose the models that best fit the data according to specific metrics [16]. Although the data set used in this study is relatively small for machine learning apps, the algorithm we used in our analysis (RF) is considered to be among the best for prediction analysis and for generating statistics of the most important predictor variables in ranking order [16]. Importantly, the RF algorithm has specific parameters that can be used to control the data set size and the imbalanced number of participants in the studied classes [16]. Furthermore, regarding prediction analysis, RF has been found to outperform traditional methods even when using relatively small data sets [16]. However, machine learning classification and regression algorithms are designed for prediction purposes and do not offer inference statistics; thus, we resorted to traditional methods such as logistic regression to obtain inference information (variable association probability metrics).

To answer RQ6, we built a multinomial logistic regression model using the SPSS software (version 28.0; IBM Corp). According to the likelihood ratio chi-square test, the full model showed a significant improvement in fit over the null model (χ^2_{36}=552.1, *P*<.001). Pearson chi-square test indicated that the model fit the data well (χ^2_{3880}=4031.4, *P*=.05), and the deviance chi-square indicated good fit (χ^2_{3880}=3490.9, *P*=.99)—indeed, in the latter 2 cases, nonsignificant test results were indicators that the model fit the data well (Field, 2018; Petrucci, 2009) [17,18].

Ethics Approval

Participants provided digital informed consent for their survey contribution. Participation was voluntary and was restricted to those aged ≥18 years. All the data were collected anonymously. In accordance with the Swiss Human Research Act (Chapter 1, Section 1, Article 2 Scope: 2c) [19], no ethics assessment was applied for as anonymously collected or anonymized health-related data do not fall under the research act's scope.

Results

Basic Descriptive Statistics on the Participants Smartphone Use

The 2 sociodemographic characteristics collected were age and sex. Age varied between 19 and 76 years (mean 45, SD 16 years) and was distributed as follows: 19 to 35 (691/1989, 34.74%); 36 to 60 (802/1989, 40.32%); and 61 to 76 years (496/1989, 24.93%). The sex distribution was as follows: male (965/1989, 48.51%); female (1006/1989, 50.57%); and nonbinary (18/1989, 0.9%). The participants' smartphone OS distribution was as follows: Android, 55.15% (1097/1989); iOS, 43.94% (874/1989); other, 0.3% (6/1989); and do not know, 0.6% (12/1989). Participants' daily average smartphone use in the 12 months preceding the study was as follows: 0.27 to 17 hours (mean 3.33, SD 2.27 h; median=3 h). Furthermore, of 1989 participants' daily average nonessential smartphone use in the 12 months preceding the study was as follows: <1 hour, 14.03% (279/1989); 1 to 3 hours, 39.22% (780/1989); 3 to 5 hours, 26.49% (527/1989); 5 to 7 hours, 9.45% (188/1989); 7 to 9 hours, 5.68% (113/1989); and >9 hours, 5.13% (102/1989). Finally, of the 1989 participants' experience of social problems owing to smartphone use in the 12 months preceding the study was as follows: definitely no, 1109 (55.76%); probably no, 557 (28%); probably yes, 262 (13.17%); and definitely yes, 61 (3.07%) respectively.

The Prevalence of Smartphone Problematic Use (RQ1)

Table 1 shows a cross tab of participant responses to the problematic smartphone use variables × the sociodemographic variables.

As shown in Table 1, 21.87% (435/1989) of the participants reported a high (4-17 h) daily average smartphone use; females and younger adults (19-35 years) were significantly more likely to be part of this group than males, adults (36-60 years), and older adults (61-76 years). A total of 46.76% (930/1989) of participants reported a high (>3 hours) daily average duration using the smartphone for nonessential activities; females and younger adults were more likely to be part of this group than males, adults, and older adults. Moreover, 16.24% (323/1989) of the participants reported having experienced social problems owing to problematic smartphone use; younger adults were more likely to be part of this group than adults and older adults. Finally, 13.42% (267/1989) of the participants reported experiencing smartphone *addiction* (scored >4 points, 1-6–point scale, in ≥4 of the 6 items in the smartphone application-based addiction scale); females and younger adults were more likely to be part of this group than males, adults, and older adults.

Table 1. Problematic smartphone use behavior in the US adult population by age category and sex (N=1989).

Demographics	DASU[a], n (%)			DADUSNA[b], n (%)		ESPDSU[c], n (%)		SA[d], n (%)	
	Low	Intermediate	High	Low	High	No	Yes	No	Yes
Total	554 (27.9)	1000 (50.3)	435 (21.9)	1059 (53.2)	930 (46.8)	1666 (83.8)	323 (16)	1722 (86.6)	267 (13.4)
Age in years									
19-35	106 (19.1)[e]	339 (33.9)[e]	246 (56.6)[e]	229 (21.6)[e]	462 (49.7)[e]	480 (28.8)[e]	211 (65.3)[e]	542 (31.4)[e]	149 (55.8)[e]
36-60	249 (44.9)[e,f]	409 (40.9)[e]	144 (33.1)[e]	456 (43.1)[e]	346 (37.2)[e]	707 (42.4)[f]	95 (29.4)[e]	705 (40.9)[e]	97 (36.3)[e]
61-76	199 (35.9)[e,g]	252 (25.2)[e]	45 (10.3)[e]	374 (35.3)[e]	122 (13.1)[e]	479 (28.8)[g]	17 (5.3)[e]	475 (27.6)[e]	21 (7.9)[e]
Sex									
Male	320 (57.8)[e]	455 (45.5)[e]	190 (43.7)[e]	552 (52.1)[e]	413 (44.4)[e]	806 (48.4)[e]	159 (49.2)[e]	843 (49)[e]	122 (45.7)[e]
Female	231 (41.7)[e]	534 (53.4)[e]	241 (55.4)[e]	501 (47.3)[e]	505 (54.3)[e]	847 (50.8)[f]	159 (49.2)[f]	864 (50.1)[f]	142 (53.2)[e]
Nonbinary	3 (0.5)[e]	11 (1.1)[e]	4 (0.9)[e]	6 (0.6)[e]	12 (1.3)[e]	13 (0.8)[g]	5 (1.5)[g]	15 (0.9)[g]	3 (1.1)[f]

[a]DASU: daily average smartphone use duration (low=0-2 h; intermediate≥2-4 h; >4-17 h).

[b]DADUSNA: daily average duration of using smartphones for nonessential activities (low=0-3 h; high≥3 h).

[c]ESPDSU: experienced social problems owing to smartphone use (in previous 12 months).

[d]SA: experiencing smartphone *addiction*; scored >4 points (on a 1-6–point scale) for ≥4 of the 6 items in the smartphone application-based addiction scale.

[e]Figures with the same exponent in each column were significantly different (P<.05). The figures with different exponents were not significantly different. For example, regarding age, 19% is significantly different from 45% and from 36%; 45% and 36% are not significantly different.

[f]Regarding age, 45% is significantly different from 19%.

[g]Regarding age, 36% is significantly different from 19%.

The Popularity of TMCSU (RQ2)

As shown in Table 2, the 3 most commonly used tools were those designed to reduce notifications (973/1989, 48.92%), reduce smartphone screen time (913/1989, 45.9%), and improve sleep time and quality (702/1989, 35.29%). Of the tools that the participants tried, the most frequently used were designed to help sleep (484/1989, 24.33%) and reduce notifications (436/1989, 21.92%), whereas the ones considered the most effective were those that removed apps from the home screen (81/291, 27.8% found it effective), deleted apps (126/574, 21.9%), and helped sleep (147/702, 20.9%). Among the parents of underage children (483/1989, 24.28%), 36.2% (175/483) targeted their children with TMCSU, with 57.1% (100/175) using them frequently and 33.1% (58/175) finding them effective.

Table 2. Participants' use of tools to monitor and control smartphone use, by tool (N=1989).

Tool category	Participants (N)	Participants using the tool, n (%)	Participants using the tool frequently[a], n (%)	Participants that considered the tool effective[a], n (%)
Tools to limit daily smartphone use duration	1989	263 (13.2)	29 (11)	39 (14.8)
Tools to reduce screen time	1989	913 (45.9)	96 (10.5)	155 (17)
Tools to calculate screen time	1989	676 (34)	96 (14.2)	115 (17)
Tools to block apps	1989	179 (9)	45 (25.1)	23 (12.8)
Tools make the smartphone less distracting	1989	115 (5.8)	39 (33.9)	17 (14.8)
Tools to improve sleep time and quality	1989	702 (35.2)	484 (68.9)	147 (20.9)
Tools to reduce notifications	1989	973 (48.9)	436 (44.8)	185 (19)
Tool to remove apps from smartphone home screen	1989	291 (14.6)	78 (26.8)	81 (27.8)
Tool to delete apps from smartphone	1989	574 (28.9)	113 (19.7)	126 (22)
Tool to control children's smartphone use	483[b]	175 (36.2)	100 (57.1)	58 (33.1)

[a]Among the number of participants using the tool.

[b]Number of participants with (<18 years) children.

Composition of the TMCSU Users' Latent Classes (RQ3 and RQ4)

Table 3 shows the composition of the TMCSU latent classes by age and sex. The first latent class (691/1989, 34.74%) was labeled *nonsmartphone-use control* (NSC) because members of this group had a low or nonexistent probability of using any of the proposed TMCSU. Males, adults, and older adults were significantly more likely to be part of this group compared with females and younger adults. The second latent class (950/1989, 47.76%) was labeled as *ineffective–smartphone-use control* (ISC) because members of this group had a moderate probability of using any of the proposed TMCSU and tended to consider that the use of these tools was ineffective. Females, younger adults, and adults were more likely to be part of this group than males and older adults.

The third latent class (348/1989, 17.49%) was labeled *effective–smartphone-use control* (ESC) because members of this group had a moderate to high probability of using most of the proposed TMCSU and tended to consider that the use of these tools was effective. Females and younger adults were more likely to be part of this group than males, adults, and older adults.

Table 3. Composition of the tools to monitor and control smartphone use user latent classes by age category and sex (N=1989).

Demographics	Latent classes, n (%)			Total participants (N=1989), n (%)
	NSC[a] (n=691)	ISC[b] (n=950)	ESC[c] (n=348)	
Age (years)				
19-35	136 (19.7)[d]	350 (36.8)[d]	205 (59)[d]	691 (34.7)
36-60	291 (42.1)[d]	399 (42)[d]	112 (32.1)[d]	802 (40.3)
61-76	264 (38.2)[d,e]	201 (21.1)[d]	31 (9)[d]	496 (25)
Sex				
Male	356 (51.5)[d,e]	454 (47.8)[d]	155 (44.5)[d,e]	965 (48.5)
Female	334 (48.3)[d,f]	487 (51.2)[d]	185 (53.1)[d,f]	1006 (50.6)
Nonbinary	1 (0.1)[d]	9 (0.9)[d]	8 (2.2)[d]	18 (0.9)

[a]NSC: nonsmartphone-use control latent class.

[b]ISC: ineffective–smartphone-use control latent class.

[c]ESC: effective–smartphone-use control latent class.

[d]Figures with the same exponent in each column were significantly different (P<.05). For example, regarding sex, 52% was significantly different from 0%; 48% was significantly different from 0%; and 52% and 48% were not significantly different.

[e]Regarding sex, 52% is significantly different from 0%.

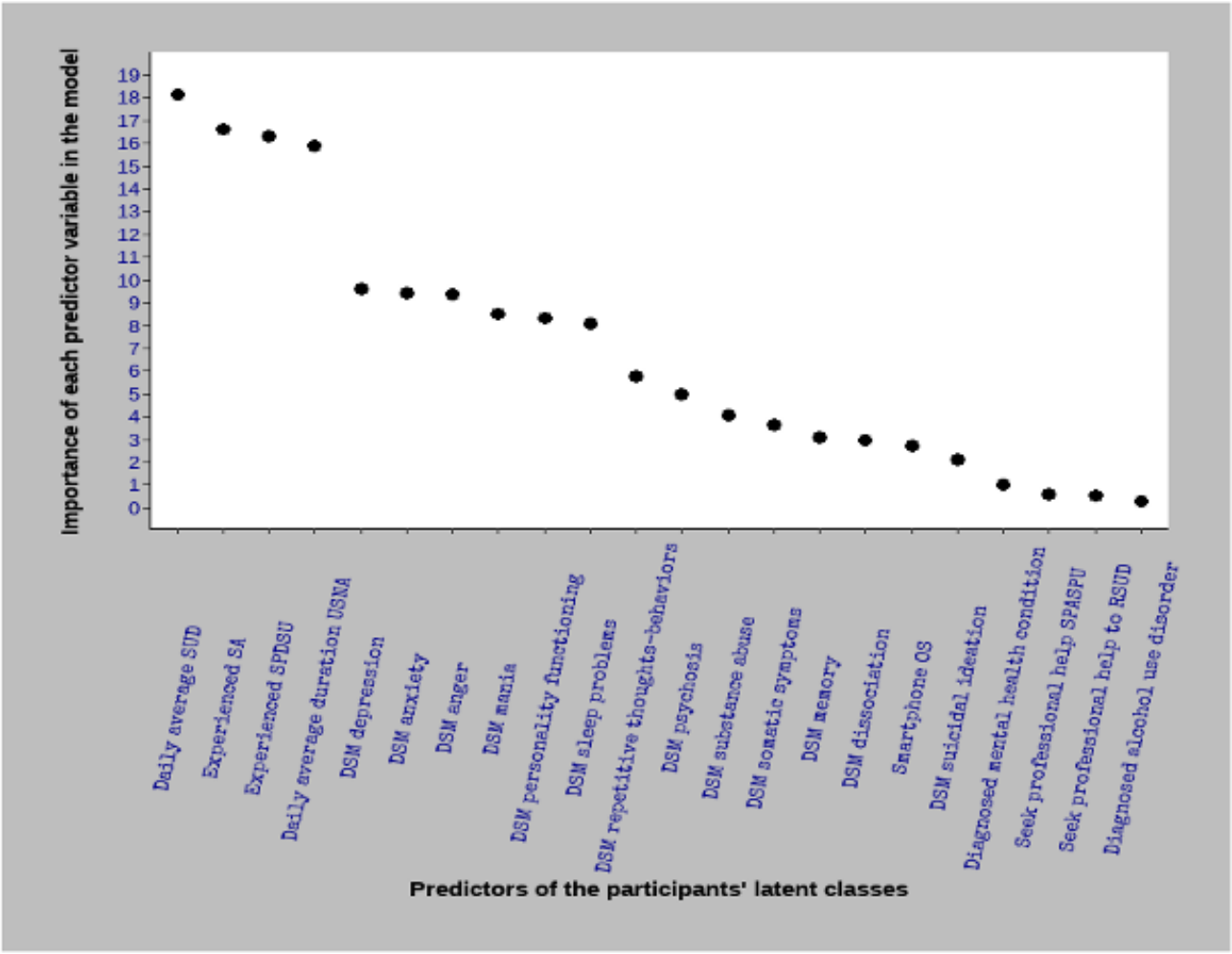
[f]Regarding sex, 48% is significantly different from 0%.

The Most Important Predictor Variables of the Latent Class Membership (RQ5)

Figure 1 shows the RF machine learning selected model MDA plot, that is, the 22 predictor variables (in decreasing order of importance) of the 3 uncovered TMCSU latent classes.

As shown in Figure 1, of the 22 predictors included in this model, the 10 most important were daily average smartphone use, experienced smartphone *addiction*, experienced social problems owing to problematic smartphone use, daily average duration of smartphone use for nonessential activities, DSM depression, DSM anxiety, DSM anger, DSM mania, DSM personality functioning, and DSM sleep problems. The 3 least important predictor variables were alcohol use disorder diagnosis, seeking professional help to reduce smartphone use, and seeking professional help because of social problems associated with smartphone use.

Figure 1. Mean decrease accuracy plot of the random forest machine learning model. It shows, in descending order of importance, the predictor variables of latent class membership. DSM: Diagnostic and Statistical Manual of Mental Disorders; OS: operating system; RSUD: reduce smartphone use duration; SA: smartphone addiction; SPASPU: social problems associated with smartphone problematic use; SPDSU: social problems owing to smartphone use; SUD: smartphone use duration; USNA: using smart phone on nonessential activities.



As discussed, the value associated with the MDA (eg, MDA=18.11) means that if the corresponding predictor variable (ie, *daily average smartphone use duration*) were removed, the model would lose that value (ie, 18.11 points) of its total accuracy score.

The importance of the RF machine learning algorithm is that it not only enables the ranking of predictor variables but also allows for variable selection and variable multicollinearity checks. Thus, only the predictor variables with an MDA of >2 and the nonmulticollinearity variables were selected for the multinomial logistic regression model presented below.

Associations Between the Covariates and the Three Uncovered Latent Classes (RQ6)

Table 4 summarizes the multinomial logistic regression models. In this model, the NSC latent class was set as the reference class and coded *0*, which means that the model was designed to predict the probability of an individual belonging to the ISC latent class (coded *1*) and the ESC latent class (coded *2*).

Table 4. Estimated β coefficients for the associations between latent classes and covariates.

Latent classes and covariates	β coefficient (SE)	OR[a] (95% CI)	*P* value
ISC[b]			
Smartphone OS[c]	−.356 (0.109)	0.7 (0.565 to 0.867)	.001[d]
Daily average smartphone use duration	.084 (0.034)	1.088 (1.017 to 1.163)	.01[d]
Dailly average duration using smartphone for nonessential activities	.162 (0.055)	1.175 (1.055 to 1.309)	.003[d]
Experienced social problems owing to problematic smartphone use	.402 (0.092)	1.495 (1.247 to 1.792)	<.001[d]
Experiencing smartphone addiction	.033 (0.012)	1.034 (1.01 to 1.059)	.006[d]
DSM[e] depression	.037 (0.089)	1.038 (0.872 to 1.234)	.68
DSM anger	.156 (0.074)	1.169 (1.011 to 1.353)	.04[d]
DSM mania	.135 (0.088)	1.144 (0.963 to 1.359)	.13
DSM anxiety	.000 (0.089)	1 (0.839 to 1.191)	.99
DSM somatic symptoms	−.025 (0.081)	0.975 (0.832 to 1.143)	.76
DSM suicidal ideation	−.296 (0.119)	0.744 (0.589 to 0.939)	.01[d]
DSM psychosis	.51 (0.243)	1.665 (1.033 to 2683)	.04[d]
DSM sleep problems	.072 (0.054)	1.074 (0.967 to 1.193)	.18
DSM memory	−0.019 (0.078)	0.981 (0.843 to 1.143)	.81
DSM repetitive thoughts-behaviors	.052 (0.124)	1.053 (0.825 to 1.344)	.68
DSM dissociation	.047 (0.108)	1.048 (0.849 to 1.295)	.66
DSM personality functioning	−.028 (0.089)	0.972 (0.816 to 1.158)	.76
DSM substance abuse	−.018 (0.086)	0.982 (0.829 to 1.164)	.84
ESC[f]			
Smartphone OS	−.321 (0.155)	0.725 (0.536 to 0.982)	.04[d]
Daily average smartphone use duration	.192 (0.042)	1.212 (1.117 to 1.316)	<.001[d]
Dailly average duration using smartphone for nonessential activities	.146 (0.076)	1.158 (0.997 to 1.344)	.06
Experienced social problems owing to problematic smartphone use	.896 (0.111)	2.45 (1.973 to 3.043)	<.001[d]
Experiencing smartphone addiction	.061 (0.017)	1.062 (1.029 to 1.097)	<.001[d]
DSM depression	.02 (0.122)	1.021 (0.804 to 1.295)	.87
DSM anger	.089 (0.100)	1.093 (0.897 to 1.331)	.38
DSM mania	.257 (0.110)	1.293 (1.042 to 1.605)	.02[d]
DSM anxiety	.035 (0.118)	1.036 (0.822 to 1.305)	.77
DSM somatic symptoms	.011 (0.107)	1.011 (0.819 to 1.247)	.92
DSM suicidal ideation	−.221 (0.140)	0.802 (0.609 to 1.056	.12
DSM psychosis	.786 (0.266)	2.195 (1.304 to 3.695)	.003[d]
DSM sleep problems	.028 (0.074)	1.029 (0.889 to 1.19)	.71
DSM memory	.101 (0.096)	1.107 (0.917 to 1.336)	.29
DSM repetitive thoughts-behaviors	−.159 (0.154)	0.853 (0.631 to 1.154)	.30
DSM dissociation	.259 (0.126)	1.296 (1.012 to 1.66)	.04[d]
DSM personality functioning	.139 (0.112)	1.149 (0.922 to 1.432)	.22
DSM substance abuse	−.321 (0.130)	0.725 (0.562 to 0.936)	.01[d]

[a]OR: odds ratio.

[b]ISC: ineffective–smartphone-use control latent class.

[c]OS: operating system.

[d]Significant at least at *P*<.05.

[e]DSM: Diagnostic and Statistical Manual of Mental Disorders.

[f]ESC: effective–smartphone-use control latent class.

In this table, the top part represents comparisons between NSC latent class, which is the reference class (baseline) and ISC latent class. The covariates significantly associated with the latent classes are the following: *DSM psychosis* (β=.51; odds ratio [OR] 1.665, 95% CI 1.033-2.683; *P*=.04); *experienced social problems owing to smartphone use* (β=.402; OR 1.495, 95% CI 1.247-1.792; *P*<.001); *daily average duration using smartphone for nonessential activities* (β=.162; OR 1.175, 95% CI 1.055-1.309; *P*=.003); *DSM anger* (β=.156; OR 1.169, 95% CI 1.011-1.353; *P*=.04); *daily average smartphone use* (β=.084; OR 1.088, 95% CI 1.017-1.163; *P*=.01); *experienced smartphone "addiction"* (β=.033; OR 1.034, 95% CI 1.01-1.059; *P*=.006); *smartphone Android OS* (β=−.356; OR 0.7, 95% CI 0.565-0.867; *P*=.001); and *DSM suicidal ideation* (β=−.296; OR 0.744, 95% CI 0.589-0.939; *P*=.01). A positive β coefficient indicates that an increase in the concerned covariate increases the probability of belonging to the ISC group, whereas a negative β coefficient indicates that an increase in the concerned covariate increases the probability of belonging to the NSC group. For the smartphone Android OS variable, for example, participants using the Android OS were significantly less likely (negative β coefficient) to be part of the ISC latent class. In the context of this analysis, the OR values can be interpreted as effect sizes. For example, if we take the DSM psychosis covariate, which has an OR of 1.665, for each unit increase in the participants' DSM psychosis score, the odds of belonging to the ISC group is 67% greater after controlling for other predictors. Note that when interpreting an OR, it is important to examine how much it deviates from 1. For instance, an OR of 0.7 means that in one group, the outcome is 30% less likely. An OR of 1.66 means that in one group, the outcome is 66% more likely. However, an OR of 2 or 3.22 means that in one group, the outcome is, respectively, 2 times or 3 times more likely.

The bottom part of the table shows comparisons between the NSC and ESC latent classes. Here, the covariates significantly associated with the latent classes are *experienced social problems owing to smartphone use* (β=.896; OR 2.450, 95% CI 1.973-3.043]; *P*<.001); *DSM psychosis* (β=.786; OR 2.195, 95% CI 1.304-3.695; *P*=.003); *DSM dissociation* (β=.259; OR 1.296, 95% CI 1.012-1.66; *P*=.04); *DSM mania* (β=.257; OR 1.293, 95% CI 1.042-1.605; *P*=.02); *daily average smartphone use* (β=.192; OR 1.212, 95% CI 1.117-1.316; *P*<.001); *experienced smartphone addiction* (β=.061; OR 1.062, 95% CI 1.029-1.097; *P*<.001); *DSM substance abuse* (β=−.321; OR 0.725, 95% CI 0.562-0.936; *P*=.01); and *smartphone Android OS* (β=−.321; OR 0.725, 95% CI 0.536-0.982; *P*=.04), this means that participants using smartphones with Android OS are significantly less likely to be part of ESC latent class.

Discussion

Principal Findings

The results suggest that females and younger adults are more likely to show high daily total smartphone use (4-17 h) and high (>3 h) daily nonessential smartphone use. Regarding smartphone *addiction*, 13.42% (267/1989) of the participants reported experiencing it, again with females and younger adults being significantly more likely to be affected. A slightly larger percentage, 16.24% (323/1989) reported social problems attributable to smartphone use, with younger adults being statistically more likely to be in this group. The higher risk among younger females has been highlighted in several previous studies [20-23] and has been linked to higher reliance on mobile phones by young females for interpersonal, social, and safety needs [24,25]. This argues for an inclusive and sex-minded design for smartphone monitoring and control tools.

Regarding the use of TMCSU, the smartphone functionalities that limit notifications and reduce screen time were the most commonly tried and used by nearly half of the sample, followed by those that improve sleep (702/1989, 35.29%). Once tried, participants were most likely to keep using sleep-related tools (484/702, 68.9%) and those that limit notifications (436/973, 44.8%). This suggests awareness of real problems such as insomnia, distractibility, and encroachment on other aspects of life caused by disruptive and excessive engagement with smartphones and is in line with increased citizen calls for more effective regulation of *Big Tech* and up-to-date legislation to curb runaway technology growth [26]. This also reflects good acceptability of these tools, suggesting that the introduction of rigorously tested and proven alternatives in the future would likely be embraced by many smartphone users.

The need for more efficacious tools is highlighted by the finding that relatively small percentages of frequent users of tools that move apps from the home screen, delete apps, and help improve sleep actually found them effective (81/291, 27.8%; 126/574, 21.9%; and 147/702, 20.9%, respectively). Similar issues were highlighted in the experiences of parents in our sample; while more than a third targeted their underage children with tools to monitor and limit their smartphone use, and more than half relied on them frequently, only a third found them effective.

The LCA revealed intriguing results. A total of 34.74% (691/1989) of the sample mapped to a class that had a low or nonexistent probability of using any queried smartphone tools, with males, adults, and older adults being significantly more likely to belong to this group than females and younger adults. This suggests that, if proven effective, the marketing of new tools that curb excessive smartphone use should focus on these subgroups. Another 17.5% (348/1989) had a moderate to high probability of using the queried tools and tended to find them

effective, with females and younger adults more likely to belong to this group than males, adults, and older adults. This suggests that females and younger adults, who in our sample were statistically more likely to experience smartphone *addiction* and to spend most of their time using their smartphone and performing nonessential smartphone activities, were also the most optimistic about the possibility of finding help on their smartphones. This is not surprising; accustomed to pursuing all activities on the web, digital natives may also gravitate toward finding help there, including for technology-mediated problems [27]. The largest latent class (965/1989, 48.52%) had a moderate probability of using the queried tools and tended to consider them ineffective, with females, younger adults, and adults being more likely to be part of this group than males and older adults.

Regarding the predictors of the extent to which participants used tools to monitor and control smartphone behaviors and whether they found them effective, no solid conclusions could be drawn. The machine learning model suggests that the most important predictors are related to smartphone use behavior, interpersonal relationships, and some psychopathological aspects (eg, daily smartphone use, smartphone *addiction*, social problems, DSM depression, DSM anxiety, DSM anger, DSM mania, DSM personality functioning, and DSM sleep), while the less important predictors are related to other psychopathological aspects (eg, alcohol use disorder diagnosis, DSM suicidal ideation, and DSM dissociation).

In addition, participants using the Android OS were more likely to not use tools to monitor or control smartphone use compared with those using iOS, perhaps suggesting inadequate marketing and outreach on the part of its maker or an inferior product or platform. Similarly, participants with high scores on the DSM suicidal ideation and substance abuse measures were more likely to not use TMCSU behavior. This may suggest a heavier reliance on smartphones among the more severely depressed or substance users, making curtailing use less appealing. Alternatively, it could suggest self-esteem– or motivation-related obstacles among participants with depression.

Other mental health conditions did not seem to discourage the use of these tools, but no clear pattern emerged as to their effectiveness. Participants with high scores on DSM mania and dissociation were statistically more likely to report effective use versus no use, possibly as a means to reduce stimulation in the former group. In contrast, those with high scores on DSM anger and total nonessential smartphone use, were more statistically likely to report ineffective use versus no use, possibly owing to the difficulty in reaching the effective threshold of tool engagement among those with heavy nonessential use or engaging appropriately with the tools among those with anger issues. In addition, some participants with high scores on *addiction*, DSM psychosis, daily average smartphone use duration, and social problems owing to use were significantly more likely to ineffectively use the TMCSU, whereas others were significantly more likely to effectively use the TMCSU.

Taken together, our data seem to portend well for the acceptability and possible effectiveness of mobile telepsychiatry help, including for conditions considered more challenging and for which digital health interventions may not have been seriously considered.

A few limitations complicate our interpretations and warrant discussion. The web-based questionnaire was based on self-reporting, which can introduce bias and compromise validity. This is true, for example, when recalling the amount of time spent, the specific tools used, and the effectiveness of the tools used. In addition, the conditions assessed—smartphone *addiction* as well as DSM-based categories—were not the product of the gold standard in-person comprehensive diagnostic evaluation and may, therefore, be unreliable. Furthermore, the sample, though large and with a broad age range and nearly equal male-female sex distribution, was exclusively US-based, potentially limiting its generalizability. The sample also included only adults aged ≥18 years, when many of the issues assessed are highly relevant to younger adolescents who are often thought to be disproportionately impacted by smartphone use and internet-related technologies. Whether our findings can be generalized to this subpopulation is unknown. In addition, the fact that our survey was exclusively on the web may also have overrepresented individuals with smartphone-related problems or those who gravitate to smartphone solutions. Furthermore, despite a study sample representative of the adult population in the United Sates across demographic variables, a selection bias related to Prolific participation or study selection by participants cannot be ruled out.

Finally, the survey was conducted during the COVID-19 pandemic, a period that witnessed heightened reliance on internet-related technologies, which likely affected participants' engagement with and perceptions of their smartphones and smartphone tools. Nevertheless, this is the first psychological evaluation of smartphone tools that curb smartphone use, and our results suggest a potentially promising future for this digital mental health intervention.

Conclusions

At >20 years of age, internet *addiction* has become a condition whose treatment is taking place over digital platforms. The old joke that asked users to "click here if you are addicted to the internet" is no longer funny insofar as users are increasingly "clicking" for that service as they seek on the web the tools and resources to address a problem that they are more aware of than ever before. Our study shows a relatively high acceptability of these tools and an openness to trying and using them, even if the effectiveness of the currently available tools remains inadequate. This is true for individuals trying to monitor or curtail their own use, as well as parents trying to achieve the same for their underage children. Given the limitations and despite 2 decades of research, of the psychopharmacological and psychotherapeutic offerings tested, the field and culture at large would benefit from rigorous scientific testing of these and other tools and their intelligent deployment with an eye toward those groups that seem most affected and those that seem most resistant. As it stands now, however, these tools are being developed, marketed, and widely adopted largely outside of any meaningful scientific scrutiny by the mental health field. This raises an important issue that the field must address: these built-in tools are often offered by the smartphone makers

themselves and as a result come with a *built-in* conflict of interest. Digital companies rely on the amount of time users spend interacting with their products for their income. Therefore, any endorsement by smartphone makers of tools that limit smartphone use should invite some skepticism, including any public relations-type motives.

Finally, the COVID-19 pandemic has brought greater acceptability to the telepsychiatry field overall, which could mean even larger adoption of smartphone tools that are meant to enhance well-being in the future. This would constitute a clear advance if these tools can be proven effective in well-designed representative research trials and suggests that the time is ripe for such research trials to be conducted.

Acknowledgments

The authors wish to acknowledge the Prolific team for their assistance in conducting the survey.

Authors' Contributions

Initial conception of the study was done by EA and YK. Concept of the survey and writing of the questions was accomplished by EA, YK, LR, FBB, RC, and RK. Recruitment procedure was carried out by RC, RK, FBB, and YK. Data analysis was performed by GVC. First draft was formulated by EA, GVC, and YK. Finally, the final draft was contributed by all the authors**.**

Conflicts of Interest

None declared.

References

1. International Statistical Classification of Diseases and Related Health Problems. 11th edition. World Health Organization. Geneva, Switzerland: World Health Organization; 2019. URL: https://icd.who.int/browse11/l-m/en [accessed 2022-04-10]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
3. Aboujaoude E. Problematic Internet use: an overview. World Psychiatry 2010 Jun;9(2):85-90 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00278.x] [Medline: 20671890]
4. Dell'Osso B, Hadley S, Allen A, Baker B, Chaplin WF, Hollander E. Escitalopram in the treatment of impulsive-compulsive internet usage disorder: an open-label trial followed by a double-blind discontinuation phase. J Clin Psychiatry 2008 Mar;69(3):452-456. [doi: 10.4088/jcp.v69n0316] [Medline: 18312057]
5. Bostwick JM, Bucci JA. Internet sex addiction treated with naltrexone. Mayo Clin Proc 2008 Feb;83(2):226-230. [doi: 10.4065/83.2.226] [Medline: 18241634]
6. Han DH, Lee YS, Na C, Ahn JY, Chung US, Daniels MA, et al. The effect of methylphenidate on Internet video game play in children with attention-deficit/hyperactivity disorder. Compr Psychiatry 2009;50(3):251-256. [doi: 10.1016/j.comppsych.2008.08.011] [Medline: 19374970]
7. Kuss DJ, Lopez-Fernandez O. Internet addiction and problematic Internet use: a systematic review of clinical research. World J Psychiatry 2016 Mar 22;6(1):143-176 [FREE Full text] [doi: 10.5498/wjp.v6.i1.143] [Medline: 27014605]
8. Aboujaoude E. Problematic Internet use two decades later: apps to wean us off apps. CNS Spectr 2019 Aug;24(4):371-373. [doi: 10.1017/S109285291800127X] [Medline: 30507371]
9. Prolific - Quickly find research participants you can trust. Prolific. URL: https://www.prolific.co/ [accessed 2022-09-29]
10. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
11. Csibi S, Griffiths MD, Cook B, Demetrovics Z, Szabo A. The psychometric properties of the Smartphone Application-Based Addiction Scale (SABAS). Int J Ment Health Addict 2018;16(2):393-403 [FREE Full text] [doi: 10.1007/s11469-017-9787-2] [Medline: 29670500]
12. Mahoney MR, Farmer C, Sinclair S, Sung S, Dehaut K, Chung JY. Utilization of the DSM-5 self-rated level 1 cross-cutting symptom measure-adult to screen healthy volunteers for research studies. Psychiatry Res 2020 Jan 27;286:112822 [FREE Full text] [doi: 10.1016/j.psychres.2020.112822] [Medline: 32086029]
13. smartphone_problematic_use_behavior / Material and data. Gitlab. URL: https://gitlab.huma-num.fr/gveracruz/smartphone_problematic_use_behavior/-/tree/main/Material%20and%20data [accessed 2022-09-29]
14. Hagenaars JA, McCutcheon AL. Applied Latent Class Analysis. Cambridge, UK: Cambridge University Press; 2002.
15. Weller BE, Bowen NK, Faubert SJ. Latent class analysis: a guide to best practice. J Black Psychol 2020 Jun 05;46(4):287-311. [doi: 10.1177/0095798420930932]
16. Breiman L. Random forests. Mach Learn 2001 Oct;45:5-32. [doi: 10.1023/A:1010933404324]
17. Field A. Discovering Statistics Using IBM SPSS Statistics. 5th edition. Thousand Oaks, CA, USA: Sage Publications; 2018.
18. Petrucci CJ. A primer for social worker researchers on how to conduct a multinomial logistic regression. J Soc Serv Res 2009 Mar 04;35(2):193-205. [doi: 10.1080/01488370802678983]

19. Federal Act on Research Involving Human Beings. Schweizerische Eidgenossenschaft. URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en [accessed 2022-09-29]
20. Al-Barashdi HS, Bouazza A, Jabur NH. Smartphone addiction among university undergraduates: a literature review. J Sci Res Rep 2015 Jan 10;4(3):210-225. [doi: 10.9734/jsrr/2015/12245]
21. Chóliz M. Mobile-phone addiction in adolescence: the test of mobile-phone dependence (TMD). Prog Health Sc 2012;2(1):33-44 [FREE Full text]
22. Kawasaki N, Tanei S, Ogata F, Burapadaja S, Loetkham C, Nakamura T, et al. Survey on cellular phone usage on students in Thailand. J Physiol Anthropol 2006 Nov;25(6):377-382 [FREE Full text] [doi: 10.2114/jpa2.25.377] [Medline: 17213690]
23. Walsh SP, White KM, Cox S, Young RM. Keeping in constant touch: the predictors of young Australians' mobile phone involvement. Comput Human Behav 2011 Jan;27(1):333-342. [doi: 10.1016/j.chb.2010.08.011]
24. Balakrishnan V, Raj RG. Exploring the relationship between urbanized Malaysian youth and their mobile phones: a quantitative approach. Telemat Inform 2012 Aug;29(3):263-272. [doi: 10.1016/j.tele.2011.11.001]
25. Pawlowska B, Potembska E. Gender and severity of symptoms of mobile phone addiction in Polish gymnasium, secondary school and university students. Curr Probl Psychiatry 2011;12(4):433-438.
26. Vogels EA. 56% of Americans support more regulation of major technology companies. Pew Research Center. 2021 Jul 20. URL: https://www.pewresearch.org/fact-tank/2021/07/20/56-of-americans-support-more-regulation-of-major-technology-companies/ [accessed 2022-09-29]
27. Aboujaoude E, Salame W. Technology at the service of pediatric mental health: review and assessment. J Pediatr 2016 Apr;171:20-24. [doi: 10.1016/j.jpeds.2015.12.009] [Medline: 26743496]

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders
ESC: effective–smartphone-use control
ISC: ineffective–smartphone-use control
LCA: latent class analysis
MDA: mean decrease accuracy
NSC: nonsmartphone-use control
OR: odds ratio
OS: operating system
RF: random forest
RQ: research question
TMCSU: tools to monitor and control smartphone use

Edited by R Kukafka; submitted 23.04.22; peer-reviewed by C Athanasopoulou, N Karnik; comments to author 31.07.22; revised version received 09.08.22; accepted 16.09.22; published 20.10.22.

Please cite as:
Aboujaoude E, Vera Cruz G, Rochat L, Courtois R, Ben Brahim F, Khan R, Khazaal Y
Assessment of the Popularity and Perceived Effectiveness of Smartphone Tools That Track and Limit Smartphone Use: Survey Study and Machine Learning Analysis
J Med Internet Res 2022;24(10):e38963
URL: https://www.jmir.org/2022/10/e38963
doi:10.2196/38963
PMID:36264627

Original Paper

Investigating the Provision and Context of Use of Hearing Aid Listening Programs From Real-world Data: Observational Study

Alessandro Pasta[1,2*], BSc, MSc, PhD; Tiberiu-Ioan Szatmari[1,2*], BSc, MSc; Jeppe Høy Christensen[3], BSc, MSc, PhD; Kasper Juul Jensen[2], BSc; Niels Henrik Pontoppidan[3], BSc, MSc, PhD; Kang Sun[1,3], BSc, MSc, PhD; Jakob Eg Larsen[1], BSc, MSc, PhD

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

[2]Department of Applied Mathematics and Computer Science, Cognitive Systems, Demant A/S, Smørum, Denmark

[3]Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

[*]these authors contributed equally

Corresponding Author:
Alessandro Pasta, BSc, MSc, PhD
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Richard Petersens Plads, Building 324
Kongens Lyngby, 2800
Denmark
Phone: 45 4525 3031
Email: alpas@dtu.dk

Abstract

Background: Listening programs enable hearing aid (HA) users to change device settings for specific listening situations and thereby personalize their listening experience. However, investigations into real-world use of such listening programs to support clinical decisions and evaluate the success of HA treatment are lacking.

Objective: We aimed to investigate the provision of listening programs among a large group of in-market HA users and the context in which the programs are typically used.

Methods: First, we analyzed how many and which programs were provided to 32,336 in-market HA users. Second, we explored 332,271 program selections from 1312 selected users to investigate the sound environments in which specific programs were used and whether such environments reflect the listening intent conveyed by the name of the used program. Our analysis was based on real-world longitudinal data logged by smartphone-connected HAs.

Results: In our sample, 57.71% (18,663/32,336) of the HA users had programs for specific listening situations, which is a higher proportion than previously reported, most likely because of the inclusion criteria. On the basis of association rule mining, we identified a primary additional listening program, *Speech in Noise*, which is frequent among users and often provided when other additional programs are also provided. We also identified 2 secondary additional programs (*Comfort* and *Music*), which are frequent among users who get ≥3 programs and usually provided in combination with *Speech in Noise*. In addition, 2 programs (*TV* and *Remote Mic*) were related to the use of external accessories and not found to be associated with other programs. On average, users selected *Speech in Noise*, *Comfort*, and *Music* in louder, noisier, and less-modulated (all $P<.01$) environments compared with the environment in which they selected the default program, *General*. The difference from the sound environment in which they selected *General* was significantly larger in the minutes following program selection than in the minutes preceding it.

Conclusions: This study provides a deeper insight into the provision of listening programs on a large scale and demonstrates that additional listening programs are used as intended and according to the sound environment conveyed by the program name.

(J Med Internet Res 2022;24(10):e36671) doi:10.2196/36671

KEYWORDS

personalized medicine; hearing aids; data logging; listening programs; sound environment; mobile phone

Introduction

Background

Untreated hearing loss is a widespread condition [1] that has repercussions at an individual [2-4] and societal level [1,5,6]. Globally, over the next 10 years, nearly 1.5 billion people can potentially benefit from having their ear and hearing problems addressed [7]. The adoption of hearing aids (HAs) has been shown to have a positive impact on the quality of life of users [8,9] and mitigate the effect of hearing loss on household income [4]. However, one of the requisites for the widespread adoption and use of HAs is user satisfaction [10]. HA users use HAs and report listening difficulties in different real-life situations, ranging from face-to-face conversations to coping with environmental sounds [11]. Therefore, to achieve high user satisfaction, HAs need to be able to cater to a wide range of situations. This is confirmed by previous research that found that one of the main reasons for not owning or not using HAs is that they do not work well in specific situations, for instance, when there is background noise [10,12,13], when listening to speech [14], or when being in a large group of people [15]. HA users can benefit from certain HA features in specific listening environments [16]. For instance, noise reduction has been found to improve noise tolerance [17] and to decrease sustained listening effort in low signal-to-noise ratio environments [18], although its impact on speech intelligibility is equivocal [17,19,20].

Therefore, programmable multimemory HAs have been introduced, which enable providing the user with multiple listening programs for specific listening situations. Currently, 41% of HA owners have such programs [21]. Listening programs set predefined rules for contextually adapting different audiological parameters such as overall gain, frequency shaping of the gain, noise reduction, and directionality. Programs can be manually selected via the HA buttons, a remote control, or a smartphone app. Users are usually advised to use a program in a specific listening situation [22]. This is reflected by the name of the program, which often conveys the situation where it is meant to be used (eg, *Speech in Noise* and *Music*) [22]. Thus, programs are a way for users to contextually adapt the device settings in specific listening situations and thereby personalize their listening experience. Therefore, investigating the use of listening programs potentially enables a deeper understanding of users' behavior and needs.

Related Work

To benefit from listening programs, HA users need to be able to characterize the listening environment adequately and actively select the appropriate program [22]. Previous research conducted on 11 experienced HA users has shown that the percentage of users who selected identical programs in the same situation (repeatability) surpassed the level corresponding to pure guess under almost all listening conditions [23]. Higher repeatability has been found in demanding listening situations [23]. These results suggest that listening programs can discernibly impact the listening experience.

Although different listening programs can potentially be beneficial and discernible for HA users, little is known about their real-world use. De Graaff et al [22] performed a scoping review on the use of multimemory devices containing several listening programs and investigated whether HA users appreciate and adequately use the option to switch between programs. Remarkably few studies were found on the use of multiple programs for various listening environments. Stelmachowicz et al [24] found that HA users did not tend to select different settings (in terms of frequency shaping of the gain) across simulated sound environments, although differences in the preferred overall gain were sometimes observed. Conversely, Keidser et al [25] found that 5 out of 27 HA users preferred different frequency response characteristics in different listening conditions, mainly in noisy environments. Similarly, Banerjee [26] found that HA users preferred the default setting most often and nondefault settings mainly in difficult listening situations. In addition, several studies found that most HA users switched between omnidirectional and directional microphone settings and that microphone preferences depend on the characteristics of the listening environment [27-30].

These studies suggest that some HA users value and use the option to switch between listening programs. However, the existing literature is sparse and dated.

While listening programs investigated in older studies used to set a constant level for an audiological parameter (eg, higher constant amount of noise reduction), nowadays listening programs set dynamic rules for contextually adapting the parameters (eg, rules that provide earlier and stronger noise reduction as the user transitions to a complex environment). However, some questions remain unanswered. First, it is not clear what motivates an HA user to obtain a multimemory HA and manually switch between programs and in which listening situations users particularly seek device personalization. Second, as highlighted in the aforementioned systematic review, little is known about the correct use of programs designated for a specific listening environment [22]. Indeed, establishing the need for a multimemory device does not guarantee that the user will immediately notice the benefits of multiple programs. The failure to match the multimemory HA settings to the communication and environmental needs of the individual may lead to delays in fully realizing its benefits [31]. None of the studies included in the systematic review examined whether a certain program was used in the correct listening environment (eg, whether users selected a *Speech in Noise* program in noisy environments) during everyday life [22].

Furthermore, most of these studies relied on self-reported measures collected over a short period. Indeed, they used diaries or questionnaires in which HA users reported use, preferences, and details of the listening environments. Whether the appropriate program is used in each listening environment cannot be derived from these data [22]. Moreover, most studies have paid little attention to the continuation of use of the listening programs after the completion of the study. On the one hand, participants might use programs during the study period but stop using them once the study finishes. On the other hand, they might need to acclimatize to the use of programs, and their preferences may only be evident after extended use [32]. In contrast to self-reported measures, data logging enables investigating the real-world behavior of a larger number of users

[33]. It allows gathering objective data about program use and objective contextual data. Moreover, it enables assessing program use with a greater temporal resolution and longitudinally making it possible to investigate detailed patterns of use, explore the long-term user behavior, and account for the acclimatization phase [34]. Investigating the use of listening programs by using objective data logging could unveil insights into how users select different listening programs under natural conditions, thereby paving the way for more personalized hearing care solutions.

Research Objective

We aimed to investigate the provision and context of use of multimemory HAs by leveraging objective data logged by smartphone-connected HAs from in-market users across several countries. First, we investigated the provision of multiple listening programs for various listening environments. Namely, we examined how many and which programs HA users have and use and whether some programs are commonly provided together. Second, we explored whether HA users use specific programs in distinct listening situations and whether such situations reflect the listening intent conveyed by the name of the program. We did so by focusing on users who repeatedly use specific programs and investigating the sound environment in which such programs are selected.

Methods

Participants and Apparatus

This study used data from a large-scale internal (Oticon A/S) database, which stores logs of HA use of HA owners who have signed up for the HearingFitness feature [35] via the Oticon ON smartphone app. The participants were the owners of Oticon Opn HAs who used the HearingFitness feature between June and September 2020. In the sign-up process, the participants actively gave their consent for data to be collected, stored, and used for research purposes on aggregated levels. No personal identifiers were collected.

Ethics Approval

No additional ethics approval was necessary for this study according to the Danish National Scientific Ethical Committee [36].

Data and Data Analysis

Using the fitting software, the hearing care professional can provide an HA user with up to 4 listening programs (ie, one for each of the 4 memory slots available in the HAs), by selecting from a list of predefined listening programs, by fine-tuning and renaming predefined listening programs, or by freely creating new ones. The hearing care professional can decide on both the quantity and the order of the provided listening programs by assigning a specific program to the preferred memory slot. In addition, when the user uses some accessories (eg, television adapter and remote microphone), the HA adds special programs on top of the 4 available memory slots.

When the HAs are connected to the smartphone, the HearingFitness feature logs time-stamped data about the interactions with the HAs, such as the selection of specific listening programs. To account for different phrasing or different languages adopted by hearing care professionals when naming the programs, similar program names were coded in fewer categories. Moreover, when the HAs were connected to the smartphone, time-stamped continuous data about the sound environment were collected every 10 minutes, and every time a listening program was selected by the user. Such data represent acoustic characteristics of the momentary sound waves sensed by calibrated HA microphones at ear level. Namely, the sound pressure level (SPL), the noise floor (NF), and the sound modulation level (SML) in decibels were measured across a broad frequency band (0.1-10 kHz) [37].

The SPL is the level output estimate from a low-pass infinite impulse response filter with a time constant of 63 milliseconds [38]. The SPL is the most used indicator of the sound wave strength and correlates well with the human perception of loudness [39]. A bottom tracker (peak detector) of the SPL is implemented with a slow dynamic attack time of 1 to 5 seconds and a fast release time of 30 milliseconds. A top tracker (valley detector) is implemented with the reverse [38]. The NF is the level of background noise in a signal and is estimated based on the bottom tracker of the SPL. The SML is derived as the difference between a top and bottom tracker of the SPL [38]. The SML describes how much the modulated variable (eg, speech) of the signal varies around its unmodulated level and can be viewed as an estimator of the temporal signal-to-noise ratio without having to separate the signal and noise.

Provision of Listening Programs

The provision of listening programs was investigated by including users who have usage information for at least 20 hours and analyzing, for each user, the programs that have been selected at least once in the 4-month period. The 20-hour threshold was adopted to ensure that the program provision was evaluated for users logging sufficient data while still including as many users as possible.

We explored the provision of listening programs by computing the number of programs provided per user and by analyzing the name and usage of the most frequently provided programs. Furthermore, we investigated the relationships between programs by determining the association rules [40] using the Apriori algorithm [41]. Such an algorithm enables exploring how 2 or more listening programs are related to one another by analyzing the programs that are frequently provided together. Given a set of n programs $P=\{p_1,p_2,\ldots,p_n\}$ and a set of users $U=\{u_1,u_2,\ldots,u_m\}$, where each user is provided with a subset of the programs in P, a rule is defined as an implication of the form $X \Rightarrow Y$, where X is the antecedent, Y is the consequent, $X,Y \subseteq P$, and $X \cap Y = \varnothing$ [41]. In determining the association rules, the default program (ie, *General*) was excluded. Indeed, the default program is available (chosen or prescribed) for nearly all users and including it in the association rules would not be of interest. Instead, the association rules related to the 5 most frequent additional listening programs were inspected. The rules were evaluated based on several metrics, including support, coverage, confidence, and lift [41]. The support of a rule defines how often the rule appears in the data set. The coverage refers to how often the antecedent of a rule appears in the data set and

measures how often the rule can be applied [42]. The confidence of a rule is defined as conf(X⇒Y)=support(X∪Y)/support(X) and can be interpreted as an estimate of the probability P(Y|X) [41], measuring how often a rule is correct out of the applicable cases. A potential issue with confidence is that an association rule having a very frequent consequent will always have high confidence. The lift addresses this concern by considering how frequent the items are in the data set. The lift of a rule is defined as lift(X⇒Y)=support(X∪Y)/(support(X)support(Y)) and can be interpreted as the deviation of the support of the whole rule from the support expected if the antecedent and the consequent were independent [41]. Finally, the likelihood of a program being provided to users with 1, 2, 3, or 4 programs was investigated. The data manipulation was performed in Python (Python Software Foundation). The association rule mining was performed in R (R Foundation for Statistical Computing) by using the *arules* package [43].

Use of Listening Programs Versus Sound Environment

Contextual program use was evaluated by analyzing the sound environment (SPL, NF, and SML) during program selection. For each logged selection of a specific listening program, the sound environment measured in a 10-minute time window centered on the program selection was considered (ie, 5 minutes preceding and 5 minutes following the selection). For each program, only users with at least 5 selections were included. Such a threshold was chosen to ensure that users' behavior was inferred from a representative sample of program selections while, at the same time, not discarding too many users. Moreover, based on the analysis described in the *Provision of Listening Programs* section, only a relevant subset of the listening programs was included.

For visualization purposes, the sound environments occurring during repeated selections of a specific program by the same user were averaged. We visually compared the distribution of users by their average sound environments occurring when selecting a specific listening program versus their average sound environments occurring when selecting the default program (ie, *General*).

Owing to the unbalanced nature of the data (unequal samples per participant, hour, etc), associations between program selections and sound environment were analyzed by using linear mixed effect (LME) models, as recommended by Oleson et al [44]. Specifically, SPL, NF, and SML were treated as dependent variables in 3 separate random intercept models defined as the following:

$$Y_{ijk} = \beta_0+\beta_1 PROGRAM_{ijk}+u_{0j}+v_{0k}+e_{ijk},$$
$$i=1,\ldots,I,\ j=1,\ldots,J,k=1,\ldots,K \quad \textbf{(1)}$$

where *i* indexes all observations (I=332,271 program selections), *j* indexes the participants (J=1312), *k* indexes the time of the day (K=24), and *Y* is the sound environment (average SPL, NF, and SML in 3 separate models) occurred in a 10-minute time window centered on program selection. The selected listening program (*PROGRAM*) was treated as fixed effect, while u_{0j} and v_{0k} are the random intercepts, respectively for the *j*-th participant and *k*-th time of day (in hours).

In addition, to account for differences in participant behavior, we fitted the data with 3 random intercept and slope models defined as the following:

$$Y_{ijk}=\beta_0+\beta_1 PROGRAM_{ijk}+u_{0j}+v_{0k}+u_{1j}PROGRAM_{ijk}+e_{ijk},$$
$$i=1,\ldots,I,\ j=1,\ldots,J,k=1,\ldots,K \quad \textbf{(2)}$$

where compared with the simpler model (equation 1), the only additional term is $u_{1j}PROGRAM_{ijk}$, where u_{1j} is the random slope varying across participants for the program effect. These relatively more complex models (equation 2) were compared with simpler models (equation 1) by conducting likelihood ratio tests.

Furthermore, we investigated whether the sound environment changed before or after the program selection by analyzing the sound environment measured in the 5 minutes preceding each program selection and the 5 minutes following it. The difference in sound environment before and after program selection was assessed by 3 separate LME models defined as the following:

$$Y_{ijk}=\beta_0+\beta_1 PROGRAM_{ijk}+\beta_2 TIMEWINDOW_{ijk}$$
$$+\beta_3 PROGRAM_{ijk}\times TIMEWINDOW_{ijk}+u_{0j}+v_{0k}+e_{ijk},$$
$$i=1,\ldots,I,\ j=1,\ldots,J,k=1,\ldots,K \quad \textbf{(3)}$$

where *i* indexes all observations (I=273,687 program selections), *j* indexes the participants (J=825), *k* indexes the time of the day (K=24), and *Y* is the sound environment (average SPL, NF, and SML in 3 separate models). The selected listening program (*PROGRAM*) and the time window (*TIMEWINDOW*, ie, *5 minutes before* or *5 minutes after*) were treated as fixed effect. The interaction between *PROGRAM* and *TIMEWINDOW* was introduced to test whether the difference in sound environment levels before and after program selection depends on which program is selected. Finally, u_{0j} and v_{0k} are the random intercepts, respectively, for the *j*-th participant and *k*-th time of day (in hours). By conducting likelihood ratio tests, these models were compared with simpler models excluding *PROGRAM*. Moreover, by conducting post hoc ANOVA tests, the significance of the variables included in equation 3 was tested. Finally, pairwise comparison (ANOVA) tests were performed on the estimated marginal means from the interaction model to test the difference in the sound environment before and after selection of each listening program.

The data manipulation and visualization were performed in Python using the NumPy [45], Pandas [46], Seaborn [47], and Scipy [48] libraries. The data analysis was performed in R using base functions, and the *lmerTest* (version 3.13 [49]) and *emmeans* (version 1.74-1 [50]) packages were used to apply LME modeling.

Results

Provision of Listening Programs

The data processing described in the *Methods* section resulted in a total of 32,336 users and 67,996 programs provided. On average, the sampled users had a connected HA use of 5.88 hours per day. However, when only considering days with at least 1 hour of connected HA use, the average connected HA use amounted to 8.81 hours per day.

Among the HA users, 57.71% (18,663/32,336) had >1 listening program (Figure 1). Almost every user (31,871/32,336, 98.56%) had the default program, *General* (Figure 1). This means that more than half of the users have at least one program for specific listening situations in addition to the default program. Furthermore, 25.8% (8344/32,336), 12.98% (4199/32,336), and 10.26% (3319/32,336) of the users had a *Speech in Noise*, *Music*, and *Comfort* program, respectively. The names of these programs convey a specific listening intent. In addition, 18.13% (5862/32,336) and 11.67% (3773/32,336) of the users had a *TV* and *Remote Mic* program, respectively. These programs are related to the use of an accessory, such as a television adapter and a remote microphone.

In addition to the provision of programs, their use was investigated by computing the percentage of time spent in each program for users with that program and at least another program. *General* was the most used program, accounting on average for 78% of the HA use time. *Speech in Noise*, *Music*, and *Comfort*, respectively, accounted for 13%, 7%, and 15% of HA use time. *TV* and *Remote Mic* accounted for 20% and 2% of HA use time, respectively.

Investigating the association rules with support ≥0.02 and confidence >0.5 (Figure 2) enables exploring the relationships between programs. In this analysis, *General* was not considered as it is uniformly provided and is not an additional listening program. The detailed metrics of the selected rules are presented in Table 1. *Speech in Noise* was not only the most common additional listening program but also a primary program that users get when also getting secondary programs. Indeed, *Speech in Noise* was the consequent of all selected rules, while either *Comfort* or *Music* was always in the antecedent set. As shown by the confidence metric in Table 1, 62.2% (2612/4199) and 71.01% (2357/3319) of the users who had either *Music* (rule 1) or *Comfort* (rule 2), respectively, also had *Speech in Noise*. Similarly, 78.76% (801/1017) of the users who had both *Music* and *Comfort* (rule 3) also had *Speech in Noise*. For these rules, the lift is >1, indicating that users are more likely to have *Speech in Noise* when they also have *Music* or *Comfort*. In contrast, although *TV* was a frequently provided program, users who had such programs were not more likely to have other listening programs.

Figure 3 confirms some of the previous findings. Almost all users have the *General* program regardless of the number of additional programs. Among the users that have 2 programs, *Speech in Noise*, *TV*, and, to a lesser extent, *Remote Mic* are more likely to be available than *Music* and *Comfort*. For users with 3 or 4 programs, the likelihood of having the primary program *Speech in Noise* grows linearly, the likelihood of having *TV* or *Remote Mic* remains relatively constant, and the likelihood of having secondary programs *Music* and *Comfort* increases.

Figure 1. Left, the number of listening programs available for each user is displayed. Right, the provision and usage of the 6 most frequently provided programs are presented. The percentage of users provided with each of the 6 programs (dark blue bars) and the percentage of usage time spent with the programs (light blue bars) are shown. The percentage of usage time spent with each program is computed for the users having that program and at least another program.

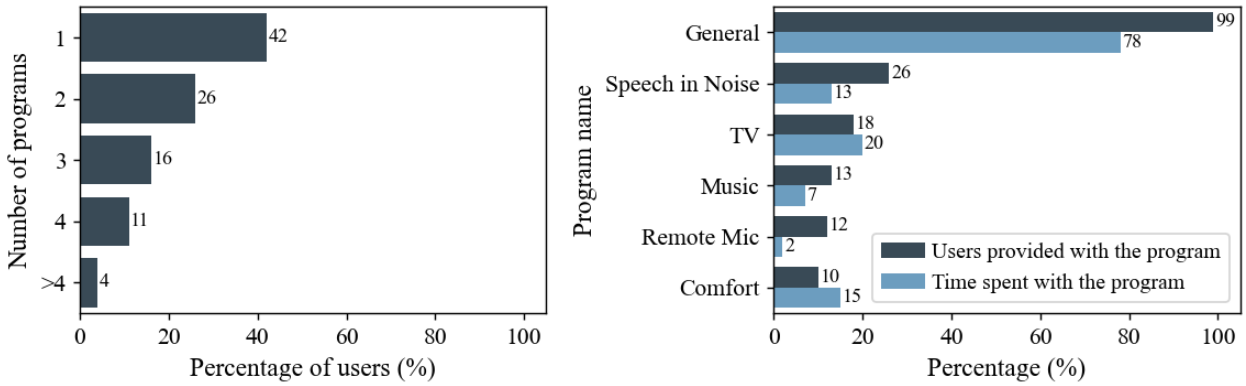


Figure 2. Association rules with support ≥0.02, confidence >0.5, and lift >1 (see the Provision of Listening Programs section). The support of each rule is indicated by the area of the circle, while the confidence is conveyed by the color intensity. *Speech in Noise* is the consequent of all rules, suggesting that it is a primary program, frequently provided when secondary programs such as *Comfort* and *Music* are also provided.

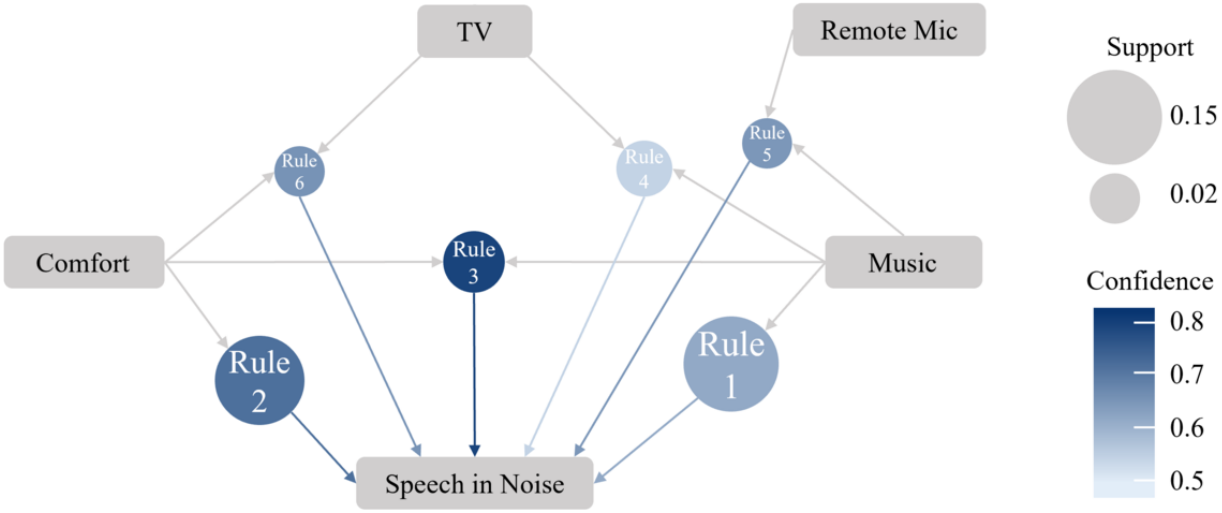
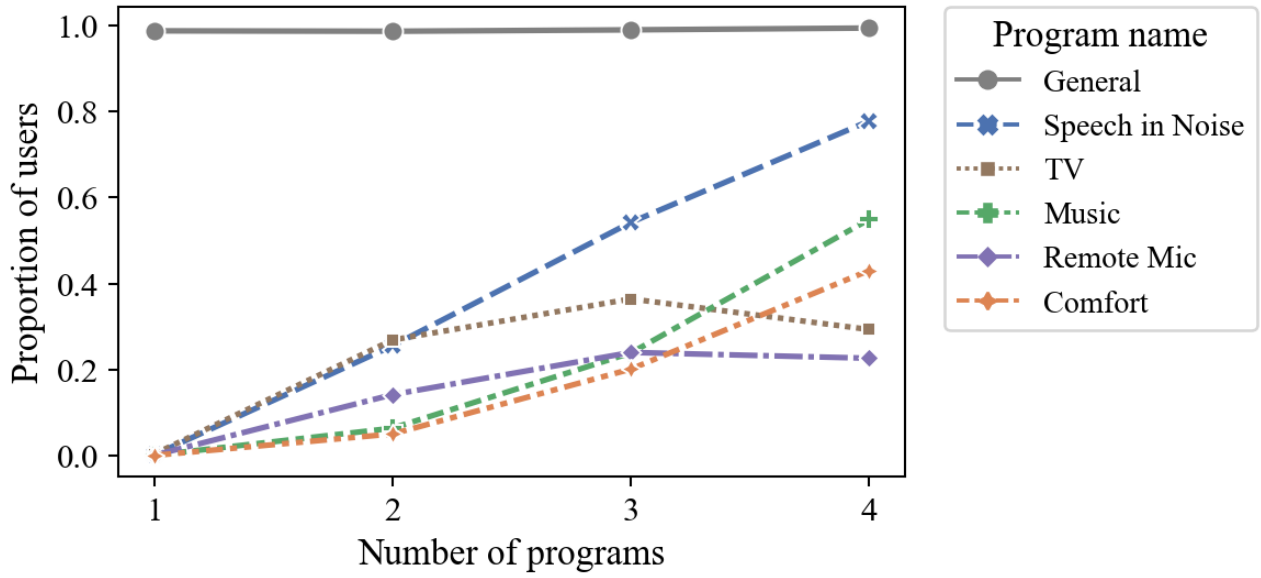


Table 1. Association rules with support ≥0.02, confidence >0.5, and lift >1.

Rule	Antecedent	Consequent	Support	Coverage	Confidence	Lift	Count
1	Music	Speech in Noise	0.14	0.23	0.62	1.35	2612
2	Comfort	Speech in Noise	0.13	0.18	0.71	1.54	2357
3	Comfort and Music	Speech in Noise	0.04	0.06	0.79	1.71	801
4	Music and TV[a]	Speech in Noise	0.03	0.05	0.56	1.21	476
5	Music and Remote Mic	Speech in Noise	0.02	0.03	0.66	1.43	401
6	Comfort and TV	Speech in Noise	0.02	0.03	0.67	1.45	377

[a]TV: television.

Figure 3. Likelihood of specific listening programs being provided to users with 1, 2, 3, or 4 programs.



Use of Listening Programs Versus Sound Environment

Since the findings presented in the *Provision of Listening Programs* section, we investigated the sound environments in which a relevant subset of the listening programs was used. We focused on programs that convey a specific listening intent, whether they are primary (*Speech in Noise*) or secondary (*Comfort* and *Music*). These 3 programs are meant to be used in specific listening situations and are not related to the use of an accessory. The data processing described in the *Methods* section resulted in a total of 332,271 program selections from 1312 users.

First, we analyzed whether the primary program (*Speech in Noise*) was selected in different listening situations compared with the default program (*General*). The upper graphs in Figure 4 display the distribution of users by their average sound environment respectively when selecting *Speech in Noise* and *General*. Users selected *Speech in Noise* in louder (higher SPL), noisier (higher NF), and less-modulated (lower SML) sound environments. Indeed, on average, users selected *Speech in Noise* at 55.1 (SD 7.4) dB SPL, 46.9 (SD 7.0) dB NF, and 17.1 (SD 4.9) dB SML, while they selected *General* at 53.0 (SD 5.6) dB SPL, 44.5 (SD 5.2) dB NF, and 18.2 (SD 3.5) dB SML. The likelihood ratio tests documented that the more complex LME models (equation 2, ie, random intercept and slope for each participant) fit the data significantly better than the simpler model (equation 1) with only a random intercept for each participant (SPL: $\chi^2_{10}=3103.7$, $P<.001$; NF: $\chi^2_{10}=4308.7$, $P<.001$; MI: $\chi^2_{10}=1806.6$, $P<.001$). The more complex model (equation 2) was fitted by setting the *General* program as the baseline condition. The coefficients of the more complex LME models (Figure 5) confirmed that *Speech in Noise* and *General* were selected in different sound environments in terms of SPL, NF, and SML (all $P<.001$). The coefficients also indicate that the scale of the difference ranges from around 0.09 to 0.19 SDs (z-score); that is, 9% to 19% of the overall SD. Moreover, inspecting individual users, the lower graphs in Figure 4 corroborate the LME outcomes and show that most of the users (614/963, 64%; 633/963, 66%; and 593/963, 62%, respectively) selected *Speech in Noise* in environments characterized by higher SPL, higher NF, and lower SML.

Second, we analyzed whether the secondary programs (*Comfort* and *Music*) were selected in specific listening situations. As shown in Figure 5, users selected both the programs in louder, noisier, and less-modulated (coefficients of LME models, all $P<.01$) sound environments compared with the sound environment in which they selected *General*. Subsequently, equation 2 was refitted by changing the contrast so that the *Speech in Noise* program represented the baseline condition. This made it possible to compare whether *Comfort* and *Music* were selected in different listening situations compared with *Speech in Noise*. *Comfort* was selected in less-loud (β=–0.029, SE 0.012, $P=.014$) and less-modulated (β=–0.031, SE 0.0125, $P=.013$) environments, whereas *Music* was selected in less-loud (β=–0.083, SE 0.011, $P<.001$) and less-noisy (β=–0.086, SE 0.0103, $P<.001$) environments.

Finally, we investigated the extent to which the sound environment changed from before to after the program selection. Figure 6 shows, for a time window near the program selection, the 5-minute running average of the difference between the sound environment when selecting a program and when selecting *General*. For all 3 programs (*Speech in Noise*, *Comfort*, and *Music*) and all 3 sound environment features (SPL, NF, and SML), a difference from *General* was observed throughout the whole 10-minute time window. In addition, the sound environment difference appeared to increase after program selection.

The likelihood ratio tests (SPL: $\chi^2_6=711.0$, $P<.001$; NF: $\chi^2_6=1597.4$, $P<.001$; and SML: $\chi^2_6=749.2$, $P<.001$) showed that the more complex models (equation 3, including *PROGRAM*, *TIMEWINDOW*, and *PROGRAM×TIME WINDOW*) fit the data significantly better than the simpler model (only including *TIMEWINDOW*).

Moreover, post hoc ANOVA tests revealed that the interaction between *PROGRAM* and *TIMEWINDOW* was significant for all 3 sound environment features (all $P<.001$). This suggests that the difference in sound environment before and after program selection depends on the specific program. The marginal effects predicted by the interaction term *PROGRAM×TIMEWINDOW* are shown in Figure 7. Pairwise comparisons (*Before* and *After*) confirmed that the sound environment gets quieter, less noisy, and more modulated (all $P<.01$) in the time window after the selection of *General* (compared with the time window before the selection). In contrast, the sound environment became louder, noisier, and less modulated (all $P<.05$) after the selection of *Speech in Noise* (compared with before the selection); noisier and less modulated (both $P<.001$) after the selection of *Comfort*; and louder and noisier (both $P<.05$) after the selection of *Music*.

Figure 4. Analysis of the sound environment (sound pressure level [SPL], noise floor [NF], and sound modulation level [SML]) in which *Speech in Noise* and *General* are selected. Compared with *General*, users select *Speech in Noise* in louder, noisier, and less-modulated environments. In the upper figures, distribution of users (using histograms and kernel density estimation) by their average sound environment when selecting *General* and *Speech in Noise*. In the lower figures, 2D histograms displaying, for each user, the sound environment when selecting *Speech in Noise* (y-axis) and *General* (x-axis). The color of the hexagon is determined by the number of users in the hexagon. The identity line (y=x) is drawn in gray. If a user experiences the same sound environment when selecting *Speech in Noise* and *General*, the corresponding hexagon falls exactly on the identity line.

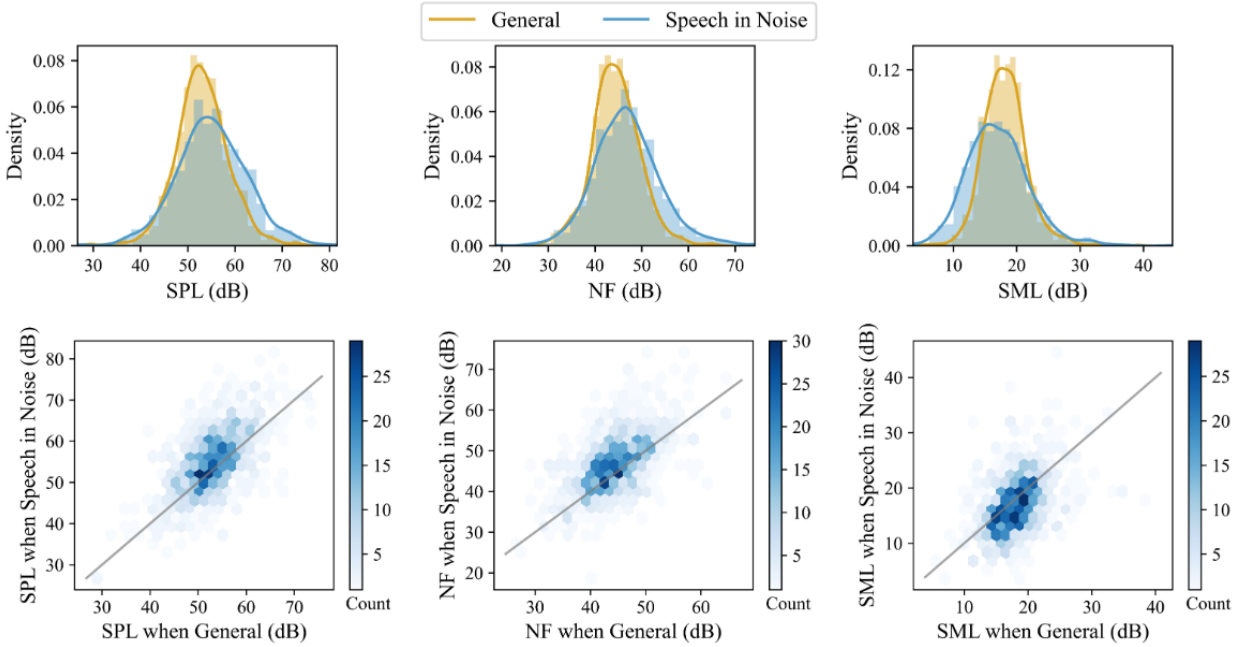


Figure 5. Coefficients and 95% CIs for predicting sound pressure level (SPL), noise floor (NF), and sound modulation level (SML) based on the selected listening program (random intercept and slope model). The baseline condition is the *General* program, so the coefficients quantify the difference in standard score between the sound environment when selecting *Speech in Noise*, *Comfort*, or *Music*, and the sound environment when selecting *General*, computed in a 10-minute interval centered on the program selection. Note that 3 separate models were fitted for predicting the 3 sound environment variables (SPL, NF, and SML). ***P*<.01; ****P*<.001.

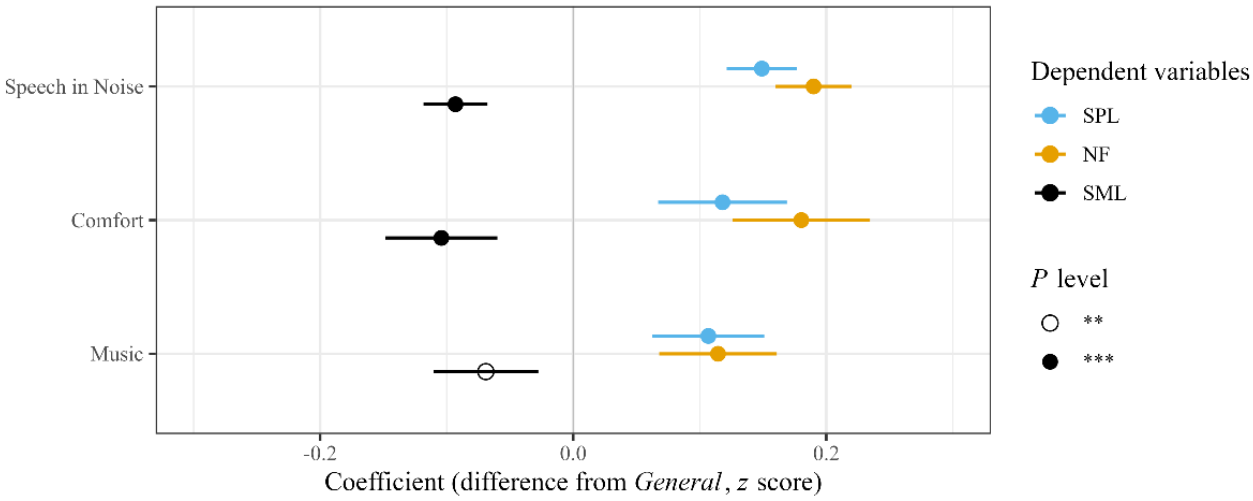


Figure 6. The 5-minute running average (SE) of the sound environment difference from *General*, computed in a time window near the program selection (ie, the solid gray line represents the sound environment when the *General* program was selected). The difference deviates from 0 throughout the whole time window. However, especially for NF and SML, the difference increases after program selection. SPL: sound pressure level; NF: noise floor; SML: sound modulation level.

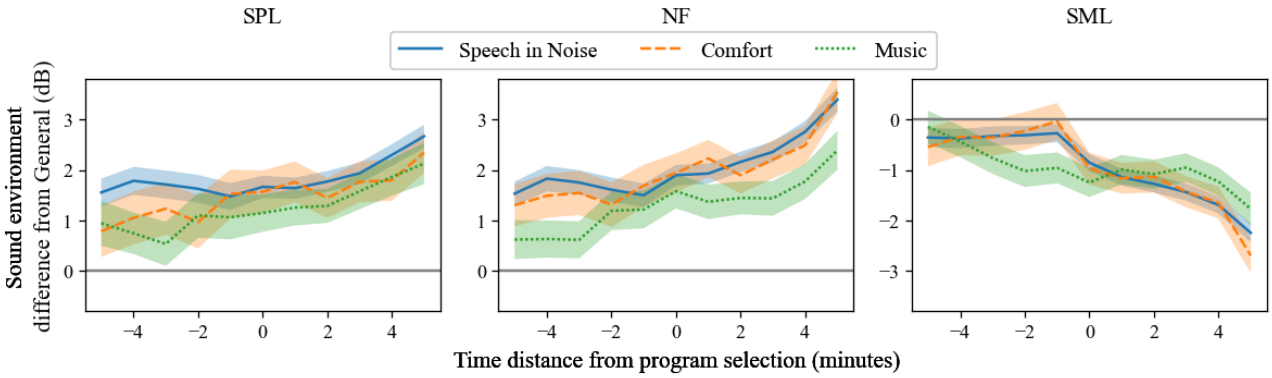
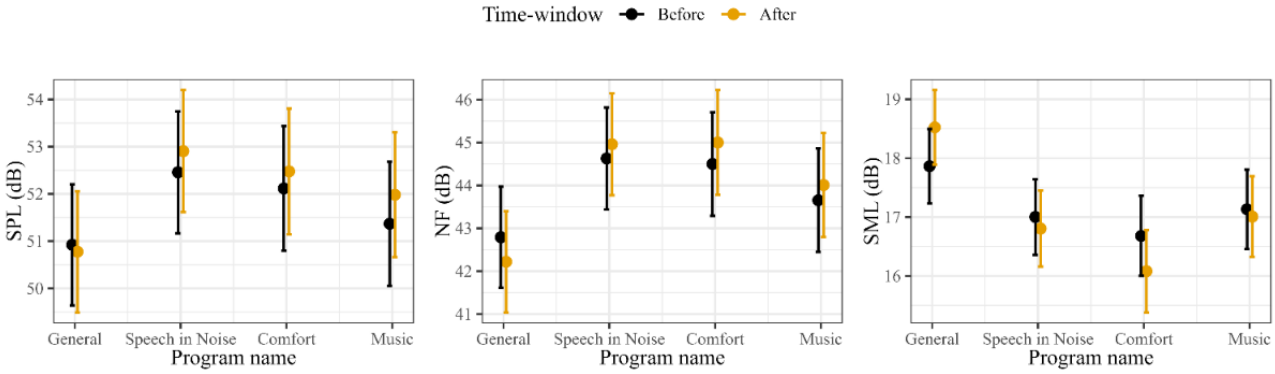


Figure 7. Predicted values of sound pressure level (SPL), noise floor (NF), and sound modulation level (SML) by selected *PROGRAM* (ie, *General*, *Speech in Noise*, *Comfort*, and *Music*) and *TIMEWINDOW* (*before*, ie, the 5-minute time window before program selection; *after*, ie, the 5-minute time window after program selection). Error bars represent the 95% CIs.



Discussion

Principal Findings

This study investigated the provision and context of use of HA listening programs by analyzing real-world data logged through smartphone-connected HAs.

Most HA users in our sample (18,663/32,336, 57.71%) were found to have listening programs for specific listening situations in addition to the *General* default program. According to a previous study analyzing self-reported data, 41% of HA owners have a program button or switch to change the HA response for different listening environments [21]. The inclusion criteria (ie, users of the HearingFitness feature via a smartphone app) and the data collection method (ie, objective data logging) of our study could explain the higher prevalence of listening programs. Among users having access to the default program and to at least one additional program, the default program was used 78% of the time. This is consistent with a previous study that estimated the default setting to be suitable 75%-85% of the time [26].

In addition to the default program, *Speech in Noise* was the most commonly provided program. By association rule mining, *Speech in Noise* was also found to be a primary additional program that users tend to get when also getting other secondary programs, such as *Comfort* and *Music*; that is, it rarely occurs that users are provided with *Comfort* and *Music* but not with *Speech in Noise*. This suggests that when users either request or are recommended additional listening programs for specific listening situations, *Speech in Noise* is provided as the primary step. This finding is consistent with previous studies reporting that HA users most frequently struggle when there is background noise [10,12,13] or when they are in a large group of people [15], and consequently, they are least likely to be satisfied with their hearing when following conversations in noise and in large groups [21]. *Comfort* and *Music* resulted to be secondary programs, frequently provided in combination with *Speech in Noise* and more likely to be provided to users having 3 or 4 programs. Similar to *Speech in Noise*, these programs signal the interest in personalizing the listening experience in a specific listening situation; that is, when it is noisy but there is no need to communicate and when listening to music. Although these situations are not as prevalent as communicating in noise, users highly motivated to personalize their experience can still benefit from adopting specific listening programs for these situations. The prevalence of the *Music* program is consistent with previous studies finding that between 30% and 67% of HA users may encounter difficulties with listening to music [51,52] and indicating that the enjoyment of listening to music with HAs could be improved by addressing problems such as distortion, acoustic feedback, insufficient or excessive gain, unbalanced frequency response, and reduced tone quality [52,53]. A

listening program dedicated to music has previously been proposed to make music more enjoyable [54].

Despite being common programs, *TV* and *Remote Mic* were provided differently than the other programs. They were frequently provided to users having only 2 programs (including *General*), but they were not frequently provided in connection with other additional programs, and their prevalence did not increase among users having >2 programs. This might be explained by the fact that such programs are related to the use of a television adapter (ie, a device that enables streaming the television sound to the HAs) or a remote microphone. Therefore, such programs show an interest in using the accessory more than in contextually adapting the HA settings through a listening program. The *TV* program was the most used program (20% of the time) besides the default program. In contrast, the *Remote Mic* program was only used 2% of the time. These findings suggest that the television adapter is extensively used by its owners, while the remote microphone is used in isolated occasions. In addition, it should be noted that selecting the *TV* program can actively modify the sound environment by either silencing the television or maintaining a normal level for other members of the household while reproducing the sound directly into the HAs. Therefore, *TV* and *Remote Mic* were not included in the sound environment analysis.

Subsequently, we analyzed the sound environment in which *Speech in Noise*, *Comfort*, and *Music* were selected. First, we found that, on average, users selected *Speech in Noise* in louder, noisier, and less-modulated environments compared with the environment in which they selected *General*. This proves that HA users select the *Speech in Noise* program in environments that possess distinct characteristics and that better resembles a conversation in noise. Second, *Comfort* was also selected in louder, noisier, and less-modulated listening environments compared with *General*, suggesting that HA users select it when they want to get relief in noisy environments. Interestingly, HA users selected *Comfort* in less-loud and less-modulated environments than when selecting *Speech in Noise*, indicating that *Comfort* is activated in situations with fewer auditory signals and likely with the intent of increasing the pleasantness of nonspecific listening. Third, *Music* was selected in louder, noisier, and less-modulated listening environments compared with *General*, but in less-loud and less-noisy environments compared with *Speech in Noise*. The music playing in the environment might explain the higher loudness and noise, although not as extreme as the *Speech in Noise* scenarios. Overall, considering that HA users are typically counseled to use a program in a specific listening situation [22], our findings suggest that they tend to follow such recommendations in the real-world use of their HAs. Moreover, the random intercept and slope model (equation 2) significantly outperformed the intercept model (equation 1), suggesting that the effect of program selection on sound environment varies among participants. Empowering users to personalize their listening experience by contextually adapting the HA settings can therefore result in more appropriate settings for some relevant listening situations.

Finally, we analyzed how the sound environment changes from a time window preceding a program selection to a time window following the program selection. For all 3 acoustic predictors, the sound environment change was different when selecting *Speech in Noise*, *Comfort*, or *Music* than when selecting *General*. Specifically, the sound environment gets louder, noisier, and less-modulated in the time window following a selection of *Speech in Noise*, while a selection of *Comfort* leads to nosier and less-modulated environments, and a selection of *Music* leads to louder and nosier environments (Figure 7).This suggests that some users tend to select additional listening programs in anticipation rather than as a reaction, to a more complex sound environment, and that the acoustic features can discriminate between them. In contrast, the sound environment gets quieter and more modulated after the selection of *General*. This indicates that some users tend to select the default program in anticipation of a less-complex sound environment. This might indicate that such users are aware of what the contextually most appropriate program is and proactively select it before entering a specific listening situation.

Limitations and Future Work

This study investigates the provision and context of use of HA listening programs by analyzing data logged by HA users who also use a smartphone app. The tech-savviness and interest in listening programs of the analyzed sample should be considered when generalizing the findings from this study. In particular, older and less–tech-savvy HA users may encounter fewer complex listening environments and therefore benefit less from multiple programs [55].

In terms of future work, it would be interesting to investigate the extent to which the provision of listening programs depends on HA users requesting a program or on the hearing care professional recommending it. Indeed, hearing care professionals traditionally have a great influence on the prescribed hearing solution, and data about the provision of listening programs might not only reflect the needs and preferences of HA users but also reflect the beliefs and knowledge of the professionals. Moreover, the role of individual predictors for the provision and use of listening programs deserves further investigation. Indeed, the benefit from a personalized and contextualized solution might depend on the degree of hearing loss or additional data characterizing the individuals such as age, prior experience with HAs, auditory cognitive capabilities, or suprathreshold hearing characteristics. Finally, the significant differences found in the sound environment occurring when using specific listening programs indicate that the analyzed sound environment features (SPL, NF, and SML) are promising candidates for predicting the selection of an additional listening program over the default program. Complementing such objective sound environment features with more subjective contextual features and with an evaluation of the listening experience (eg, via an ecological momentary assessment) might also enable a deeper understanding of the provision and use of HA listening programs.

Authors' Contributions

AP and TIS conceived and designed the study, organized the database, and performed part of the data analysis. AP wrote the manuscript. JHC performed part of the data analysis. JEL, KJJ, NHP, and KS supervised the findings and revised the final manuscript. All authors contributed to the manuscript and read and approved the submitted version.

Conflicts of Interest

AP, TIS, and KJJ are employed by Demant A/S. JHC, NHP, and KS are employed by Oticon A/S.

References

1. Deafness and hearing loss. World Health Organization. 2021 Jan 4. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2021-05-25]
2. Jayakody D, Friedland P, Eikelboom R, Martins R, Sohrabi H. A novel study on association between untreated hearing loss and cognitive functions of older adults: baseline non-verbal cognitive assessment results. Clin Otolaryngol 2018 Feb;43(1):182-191. [doi: 10.1111/coa.12937] [Medline: 28710824]
3. Kirkwood DH. Major survey documents negative impact of untreated hearing loss on quality of life. Hearing J 1999;52(7):32. [doi: 10.1097/00025572-199907000-00004]
4. Kochkin S. The impact of untreated hearing loss on household income. Better Hearing Institute. 2005 Aug. URL: http://ihsinfo.org/IhsV2/News_Info/MarkeTrak7_ImpactUntreatedHLIncome.pdf [accessed 2021-05-24]
5. Reed NS, Altan A, Deal JA, Yeh C, Kravetz AD, Wallhagen M, et al. Trends in health care costs and utilization associated with untreated hearing loss over 10 years. JAMA Otolaryngol Head Neck Surg 2019 Jan 01;145(1):27-34 [FREE Full text] [doi: 10.1001/jamaoto.2018.2875] [Medline: 30419131]
6. McDaid D, Park A, Chadha S. Estimating the global costs of hearing loss. Int J Audiol 2021 Mar;60(3):162-170. [doi: 10.1080/14992027.2021.1883197] [Medline: 33590787]
7. World Report on Hearing. Geneva: World Health Organization; Mar 3, 2021. URL: https://www.who.int/publications/i/item/world-report-on-hearing
8. Kochkin S, Rogin C. Quantifying the obvious: the impact of hearing instruments on quality of life. Hear Rev 2000;7(1):6-34 [FREE Full text]
9. Vuorialho A, Karinen P, Sorri M. Effect of hearing aids on hearing disability and quality of life in the elderly. Int J Audiol 2006 Jul;45(7):400-405. [doi: 10.1080/14992020600625007] [Medline: 16938798]
10. Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. Int J Audiol 2009 Apr;48(4):183-195. [doi: 10.1080/14992020802572627] [Medline: 19363719]
11. Galvez G, Turbin M, Thielman E, Istvan J, Andrews J, Henry J. Feasibility of ecological momentary assessment of hearing difficulties encountered by hearing aid users. Ear Hear 2012;33(4):497-507 [FREE Full text] [doi: 10.1097/AUD.0b013e3182498c41] [Medline: 22531573]
12. Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of hearing aids and assistive listening devices in an older Australian population. J Am Acad Audiol 2010;21(10):642-653. [doi: 10.3766/jaaa.21.10.4] [Medline: 21376005]
13. Aazh H, Prasher D, Nanchahal K, Moore B. Hearing-aid use and its determinants in the UK National Health Service: a cross-sectional study at the Royal Surrey County Hospital. Int J Audiol 2015 Mar;54(3):152-161. [doi: 10.3109/14992027.2014.967367] [Medline: 25395258]
14. Takahashi G, Martinez C, Beamer S, Bridges J, Noffsinger D, Sugiura K, et al. Subjective measures of hearing aid benefit and satisfaction in the NIDCD/VA follow-up study. J Am Acad Audiol 2007 Apr;18(4):323-349. [doi: 10.3766/jaaa.18.4.6] [Medline: 17580727]
15. Gallagher N, Woodside JV. Factors affecting hearing aid adoption and use: a qualitative study. J Am Acad Audiol 2018 Apr;29(4):300-312. [doi: 10.3766/jaaa.16148] [Medline: 29664724]
16. Pasta A, Petersen MK, Jensen KJ, Pontoppidan NH, Larsen JE, Christensen JH. Measuring and modeling context-dependent preferences for hearing aid settings. User Model User Adap Inter 2022 May 12. [doi: 10.1007/s11257-022-09324-z]
17. Brons I, Houben R, Dreschler W. Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. Trends Hear 2014 Oct 13;18:2331216514553924 [FREE Full text] [doi: 10.1177/2331216514553924] [Medline: 25315377]
18. Fiedler L, Seifi Ala T, Graversen C, Alickovic E, Lunner T, Wendt D. Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise-a combined pupillometry and EEG study. Ear Hear 2021;42(6):1590-1601. [doi: 10.1097/AUD.0000000000001050] [Medline: 33950865]
19. Wong L, Chen Y, Wang Q, Kuehnel V. Efficacy of a hearing aid noise reduction function. Trends Hear 2018;22:2331216518782839 [FREE Full text] [doi: 10.1177/2331216518782839] [Medline: 29956591]
20. Kim HJ, Lee JH, Shim HJ. Effect of digital noise reduction of hearing aids on music and speech perception. J Audiol Otol 2020 Oct;24(4):180-190 [FREE Full text] [doi: 10.7874/jao.2020.00031] [Medline: 32575952]
21. Picou E. MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids. Semin Hear 2020 Feb;41(1):21-36 [FREE Full text] [doi: 10.1055/s-0040-1701243] [Medline: 32047346]

22. de Graaff F, Huysmans E, Ket JC, Merkus P, Goverts ST, Leemans CR, et al. Is there evidence for the added value and correct use of manual and automatically switching multimemory hearing devices? A scoping review. Int J Audiol 2018 Mar;57(3):176-183. [doi: 10.1080/14992027.2017.1385864] [Medline: 29017358]
23. Berninger E, Nordström I. Multiple memory hearing aid. Consistency of program-usage in real-world listening situations. Scand Audiol 1997;26(4):252-256. [doi: 10.3109/01050399709048016] [Medline: 9428035]
24. Stelmachowicz P, Lewis D, Carney E. Preferred hearing-aid frequency responses in simulated listening environments. J Speech Hear Res 1994 Jun;37(3):712-719. [doi: 10.1044/jshr.3703.712] [Medline: 8084201]
25. Keidser G, Pellegrino A, Delifotis A, Ridgway J, Clarke M. The use of different frequency response characteristics in everyday environments. Aust J Audiol 1997;19(1):9-21.
26. Banerjee S. Hearing aids in the real world: use of multimemory and volume controls. J Am Acad Audiol 2011 Jun 06;22(6):359-374. [doi: 10.3766/jaaa.22.6.5] [Medline: 21864473]
27. Cord M, Surr R, Walden B, Olson L. Performance of directional microphone hearing aids in everyday life. J Am Acad Audiol 2002 Jun;13(6):295-307. [Medline: 12141387]
28. Walden B, Surr R, Cord M, Dyrlund O. Predicting hearing aid microphone preference in everyday listening. J Am Acad Audiol 2004 May;15(5):365-396. [doi: 10.3766/jaaa.15.5.4] [Medline: 15506498]
29. Palmer C, Bentler R, Mueller G. Evaluation of a second-order directional microphone hearing aid: II. Self-report outcomes. J Am Acad Audiol 2006 Mar;17(3):190-201. [doi: 10.3766/jaaa.17.3.5] [Medline: 16646279]
30. Ricketts T, Henry P, Gnewikow D. Full time directional versus user selectable microphone modes in hearing aids. Ear Hear 2003 Oct;24(5):424-439. [doi: 10.1097/01.AUD.0000094555.89110.0A] [Medline: 14534412]
31. Kuk FK. Clinical considerations in fitting a multimemory hearing aid. Am J Audiol 1993 Nov 01;2(3):23-27. [doi: 10.1044/1059-0889.0203.23] [Medline: 26661426]
32. Kuk F. Evaluation of the efficacy of a multimemory hearing aid. J Am Acad Audiol 1992 Sep;3(5):338-348. [Medline: 1421470]
33. Pasta A, Szatmari T, Christensen J, Jensen K, Pontoppidan N, Sun K, et al. Clustering users based on hearing aid use: an exploratory analysis of real-world data. Front Digit Health 2021;3:725130 [FREE Full text] [doi: 10.3389/fdgth.2021.725130] [Medline: 34713197]
34. Wright D, Gagné JP. Acclimatization to hearing aids by older adults. Ear Hear 2021;42(1):193-205. [doi: 10.1097/AUD.0000000000000913] [Medline: 32769437]
35. Laplante-Lévesque A, Pontoppidan N, Mazevski A, Schum D, Behrens T, Porsbo M. Data-driven Hearing Care With HearingFitnessTM: Oticon Shares Openly Its Innovative Vision and Roadmap. Copenhagen, Denmark: Oticon A/S; 2017.
36. What should I report? National Scientific Ethics Committee. 2020 Jul 10. URL: https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde [accessed 2021-05-31]
37. Christensen JH, Pontoppidan NH, Rossing R, Anisetti M, Bamiou D, Spanoudakis G, et al. Fully synthetic longitudinal real-world data from hearing aid wearers for public health policy modeling. Front Neurosci 2019 Aug 13;13:850 [FREE Full text] [doi: 10.3389/fnins.2019.00850] [Medline: 31456658]
38. Christensen J, Saunders G, Porsbo M, Pontoppidan N. The everyday acoustic environment and its association with human heart rate: evidence from real-world data logging with hearing aids and wearables. R Soc Open Sci 2021 Feb 17;8(2):201345 [FREE Full text] [doi: 10.1098/rsos.201345] [Medline: 33972852]
39. Long M. Fundamentals of acoustics. In: Architectural Acoustics (Second Edition). Cambridge, Massachusetts, United States: Academic Press; 2014.
40. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Rec 1993 Jun;22(2):207-216. [doi: 10.1145/170036.170072]
41. Hahsler M, Grün B, Hornik K. arules - A computational environment for mining association rules and frequent item sets. J Stat Softw 2005;14(15):1-25. [doi: 10.18637/jss.v014.i15]
42. coverage: Calculate coverage for rules. rrdr.io. 2022. URL: https://rdrr.io/cran/arules/man/coverage.html [accessed 2021-04-06]
43. Hahsler M, Buchta C, Gruen B, Hornik K. arules: mining association rules and frequent itemsets. CRAN. 2021. URL: https://cran.r-project.org/web/packages/arules/index.html [accessed 2021-09-12]
44. Oleson J, Jones M, Jorgensen E, Wu YH. Statistical considerations for analyzing ecological momentary assessment data. J Speech Lang Hear Res 2022 Jan 12;65(1):344-360 [FREE Full text] [doi: 10.1044/2021_JSLHR-21-00081] [Medline: 34910571]
45. Harris C, Millman K, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020 Sep;585(7825):357-362 [FREE Full text] [doi: 10.1038/s41586-020-2649-2] [Medline: 32939066]
46. The pandas development team. pandas-dev/pandas: Pandas. Zenodo. 2020 Feb. URL: https://pandas.pydata.org/about/citing.html [accessed 2022-09-19]
47. Waskom M. seaborn: statistical data visualization. J Open Source Softw 2021 Apr;6(60):3021. [doi: 10.21105/joss.03021]
48. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar 3;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]

49. Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017 Dec 6;82(13):1-26. [doi: 10.18637/jss.v082.i13]
50. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. The Comprehensive R Archive Network. 2022 Sep 8. URL: https://cran.r-project.org/web/packages/emmeans/index.html [accessed 2022-09-20]
51. Leek M, Molis M, Kubli L, Tufts J. Enjoyment of music by elderly hearing-impaired listeners. J Am Acad Audiol 2008 Jun;19(6):519-526. [doi: 10.3766/jaaa.19.6.7] [Medline: 19253784]
52. Greasley A, Crook H, Fulford R. Music listening and hearing aids: perspectives from audiologists and their patients. Int J Audiol 2020 Sep;59(9):694-706. [doi: 10.1080/14992027.2020.1762126] [Medline: 32615052]
53. Madsen S, Moore B. Music and hearing aids. Trends Hear 2014 Oct 31;18:2331216514558271 [FREE Full text] [doi: 10.1177/2331216514558271] [Medline: 25361601]
54. Hockley N, Bahlmann F, Chasin M. Programming hearing instruments to make live music more enjoyable. Hear J 2010 Sep;63(9):30. [doi: 10.1097/01.HJ.0000388538.27405.93]
55. Klein K, Wu Y, Stangl E, Bentler R. Using a digital language processor to quantify the auditory environment and the effect of hearing aids for adults with hearing loss. J Am Acad Audiol 2018 Apr;29(4):279-291 [FREE Full text] [doi: 10.3766/jaaa.16114] [Medline: 29664722]

Abbreviations

HA: hearing aid
LME: linear mixed effect
NF: noise floor
SML: sound modulation level
SPL: sound pressure level

Edited by G Eysenbach; submitted 21.01.22; peer-reviewed by E Jorgensen, T Lefèvre; comments to author 18.02.22; revised version received 13.07.22; accepted 28.07.22; published 17.10.22.

Please cite as:
Pasta A, Szatmari TI, Christensen JH, Jensen KJ, Pontoppidan NH, Sun K, Larsen JE
Investigating the Provision and Context of Use of Hearing Aid Listening Programs From Real-world Data: Observational Study
J Med Internet Res 2022;24(10):e36671
URL: https://www.jmir.org/2022/10/e36671
doi:10.2196/36671
PMID:36251349

Original Paper

Differentiated Effects and Determinants of Home Blood Pressure Telemonitoring: Three-Year Cohort Study in Jieshou, Anhui, China

Qun Xue[1*], BSc; Xuewu Zhang[2*], MD; Rong Liu[1], MSc; Xiaoqin Guan[1], BSc; Guocheng Li[1], BSc; Linhai Zhao[1], MSc; Qian Wang[2], BSc; Debin Wang[1], PhD; Xingrong Shen[1], PhD

[1]School of Health Service Management, Anhui Medical University, Hefei, China

[2]Department of Public Health, Jieshou Hospital, Fuyang, China

[*]these authors contributed equally

Corresponding Author:
Xingrong Shen, PhD
School of Health Service Management
Anhui Medical University
81 Meishan Road
Hefei, 230032
China
Phone: 86 13505612172
Email: shenxr@ahmu.edu.cn

Abstract

Background: Home blood pressure telemonitoring (HBPT) is witnessing rapid diffusion worldwide. Contemporary studies documented mainly short-term (6-12 months) effects of HBPT, and there are limited data about its uptake.

Objective: The aim of this study was to explore the 3-year use and determinants of HBPT, and the interactions with systolic and diastolic blood pressure (SBP/DBP) and overall blood pressure (BP) control rate.

Methods: HBPT records were obtained from a 3-year cohort of 5658 patients with hypertension in Jieshou, Anhui, China, and data from a structured household survey of a random sample (n=3005) of the cohort. The data analysis comprised (1) timeline trajectories of the rates of monthly active HBPT and mean SBP/DBP for overall and subgroups of patients with varied start-month SBP/DBP; and (2) multivariable linear, logistic, and percentile regression analyses using SBP/DBP, BP control rate, and yearly times of HBPT as the dependent variable, respectively.

Results: HBPT was followed by mixed changes in mean monthly SBP/DBP for varied patient groups. The magnitude of changes ranged from –43 to +39 mmHg for SBP and from –27 to +15 mmHg for DBP. The monthly rates of active HBPT all exhibited a rapid and then gradually slower decline. When controlled for commonly reported confounders, times of HBPT in the last year were found to have decreasing correlation coefficients for SBP/DBP (from 0.16 to –0.35 and from 0.11 to –0.35, respectively) and for BP control rate (from 0.53 to –0.62).

Conclusions: HBPT had major and "target-converging" effects on SBP/DBP. The magnitude of changes was much greater than commonly reported. BP, variation in BP, and time were the most important determinants of HBPT uptake. Age, education, duration of hypertension, family history, and diagnosis of hypertension complications were also linked to the uptake but at weaker strength. There is a clear need for differentiated thinking over the application and assessment of HBPT, and for identifying and correcting/leveraging potential outdated/new opportunities or beliefs.

(J Med Internet Res 2022;24(10):e37648) doi:10.2196/37648

KEYWORDS

blood pressure; home telemonitoring; effect; influence factors; China

Introduction

Home blood pressure telemonitoring (HBPT) is recommended in current hypertension management guidelines, and is witnessing rapid diffusion worldwide [1-3]. Various randomized controlled trials (RCTs) have documented marginal to moderate effects of HBPT on blood pressure (BP), ranging from a 3 to 8 mmHg reduction in systolic blood pressure (SBP) and a 1 to 4 mmHg reduction in diastolic blood pressure (DBP) [4-6]. Studies also reported changes following HBPT in terms of quality of

life, risk of cardiovascular complications, and costs due to hypertension-related service use and other outcome measures [7,8]. These effects are attributed mainly to "BP-guided" use of professional care and self-management, including self-titration of and compliance with antihypertensive medication [9,10].

Given fluctuating BP readings; changing stages (eg, normal BP, high-normal BP, grades 1 and 2 hypertension) [11] and type of hypertension (eg, office or "white coat" hypertension, masked hypertension, isolated systolic hypertension); and the varied physical, psychological, and socioeconomic conditions of patients, the actual effects of HBPT may differ greatly from patient to patient and according to the time of measurement. However, published studies on HBPT have generally adopted a "nondifferentiated" approach, focusing primarily on comparing the effects in the intervention group as a whole with those in the control group as a whole [12-14]. Although a small number of RCTs documented BP reductions for specific subgroups such as patients with inadequate baseline BP control [6,15], little is known about whether and how the effects and determinants of HBPT differ across patient groups with a varied level/stage of BP. Despite indications that the greatest effect of HBPT on BP control is usually achieved in the first months of the intervention, this is based on studies with a relatively short duration (less than 1 year) and its sustainability over the long term remains to be proven [16-18].

China has witnessed a rapid increase in the use of HBPT over the past decade. More and more residents are buying and using various types of HBPT devices. However, there is a general paucity of data about the effects and determinants of HBPT. Similar to studies in other countries, the limited publications on HBPT in China have focused primarily on comparing BP differences between the intervention and control groups, with little attention being paid to the determinants and differentiated effects of HBPT.

To fill this gap, the aim of this study was to use data from a relatively large-scale (5658 patients with hypertension) and long-term (up to 40 months) cohort in Jieshou, Anhui, China, for performing a relatively in-depth analysis of HBPT, with particular attention placed on comparing its effects and determinants across patient groups with varied levels of BP. As an inland county located in the middle and east of China, Jieshou is representative of the majority of counties in the nation.

Methods

Study Sites and Subjects

The study was built upon two related and ongoing projects. The first was initiated by Jieshou Hospital, Anhui province, China, which aimed to improve hypertension management via HBPT. The project covered all patients diagnosed with hypertension (N=5658) in all villages (N=48) served by the Jieshou Hospital Consortium. The HBPT involved an electronic oscillometric upper-arm BP monitor installed with a voice speaker capable of automatically stating the resultant measurements and educational messages to the patient. The monitors were provided by IFLYTEK Co Ltd, and were confirmed to be easily useable by ordinary residents. The readings of the HBPT were synchronously sent to a remote central data center.

The second project is an RCT registered in ISRCTN (10999269). This project used a cluster randomized sample (n=3005) of the participants in the above HBPT project to test the efficacy of a novel personalized hypertension management package [19].

By the time this study was carried out, the HBPT project had gathered BP readings from the participants for over 40 months and the RCT had completed the baseline assessment, including a structured baseline household survey.

Data Content and Collection

This study used the records from the HBPT project described above and part of the data from the corresponding baseline household survey. Each HBPT record consisted of four items: SBP, DBP, pulse per minute, and measurement date and time. The household survey took place from April to July 2021 via a structured questionnaire administered face to face. This study used 24 items from the questionnaire, soliciting information about: (1) sociodemographic characteristics, including age, sex, and education; (2) body height and weight; (3) age when hypertension was first diagnosed; and (4) hypertension-related symptoms and diagnoses (Multimedia Appendix 1).

Data Processing and Analysis

Data analysis comprised three components: (1) descriptive statistics (numbers and percentages) of study subjects by sociodemographic categories, (2) calculation and presentation (in trajectory lines) of the rates of monthly active HBPT for overall subjects and for subgroups with varied mean SBP/DBP in the first month, (3) multivariable linear and percentile regression modeling of times of HBPT and SBP/DBP in the last year, and (4) multivariable logistic regression modeling of BP control rate.

The rate of monthly active HBPT was defined as the proportion of patients who had performed HBPT at least one time in the month under concern. The multivariable linear, logistic, and percentile regression models used similar independent, exposure, and confounder variables. The dependent variables included times of HBPT in the past year for overall participants and subgroups with varied mean SBP/DPB from HBPT in the last year and the BP control rate in the last year. The exposure variables consisted of mean SBP/DBP and variations in the coefficients of SBP/DBP in the last year. The confounder variables comprised sociodemographics and health conditions. The monthly mean SBP/DBP of any patient was defined as their hourly mean SBP/DBP, calculated as the sum of all SBP/DBP readings recorded within a given hour (eg, 8:00-8:59 AM), multiplied by the number of records within the same hour. The BP control rate was computed as the times of BP readings meeting SBP<140 mmHg and DBP<90 mmHg in the past year multiplied by the total BP readings during the same period.

The analysis regarding the monthly active HBPT used all participants enrolled in the HBPT project, whereas the regression modeling used all of the participants involved in the baseline survey. The logarithm of times using HBPT in the last year was used to transform the variable into a normal distribution.

Detailed value assignment is shown in Multimedia Appendix 1. All quantitative and ordinal variables were standardized using Z-scores before the multivariable regression modeling.

Ethics Approval

This study has been approved by Anhui Medical University Biomedical Ethics Committee (number 20200936) and all the participants have signed (for those who are literate) or ticked (for those who are illiterate) the consent form.

Results

Sociodemographics of Study Participants

Of the 3005 participants recruited in the baseline survey, 57% were women. The average age of the participants was 65.50 years. Their duration of hypertension was 9.50 years on average. Over half of the respondents had a family history of hypertension (Table 1). Although detailed data about BMI and hypertension-related symptoms and diagnoses were not available for the 5658 participants in the HBPT project, they shared compatible sociodemographics with the above 3005 survey participants since the latter were a randomized sample of the former. High-normal BP formed the bulk type of hypertension (130≤SBP≤139 mmHg and/or 85≤DBP≤89 mmHg, 43.09%), followed by Grade 1 hypertension (140≤SBP≤159 mmHg and/or 90≤DBP≤99 mmHg, 32.21%) and normal BP (SBP<130 mmHg and DBP<85 mmHg, 21.37%).

Table 1. Sociodemographic and hypertension-related characteristics of participants (N=3005).

Variables	Sex		Total, n (%)
	Male, n (%)	Female, n (%)	
Age (years)			
≤50	93 (7.20)	102 (5.95)	195 (6.49)
51-60	345 (26.72)	506 (29.52)	851 (28.32)
61-70	404 (31.29)	500 (29.17)	904 (30.08)
>70	449 (34.78)	606 (35.36)	1055 (35.11)
Education			
No school education	234 (18.14)	1037 (60.61)	1271 (42.35)
Primary school	411 (31.86)	518 (30.27)	929 (31.96)
Middle school or higher	645 (50.00)	156 (9.12)	801 (26.69)
BMI			
<18.5	13 (1.06)	16 (0.97)	29 (1.00)
1.8.5-23.9	307 (24.92)	390 (23.58)	697 (24.15)
24-27.9	510 (41.40)	698 (42.20)	1208 (41.86)
≥28	402 (32.63)	550 (33.25)	952 (32.99)
Duration of hypertension (years)			
≤4	401 (31.40)	484 (28.69)	885 (29.86)
5-8	320 (25.06)	433 (25.67)	753 (25.40)
9-12	252 (19.73)	315 (18.67)	567 (19.13)
>12	304 (23.81)	455 (26.97)	759 (25.61)
Family history of hypertension			
Yes	642 (54.64)	808 (51.50)	1450 (52.84)
No	533 (45.36)	761 (48.50)	1294 (47.16)
Number of hypertension-related symptoms			
≤4	670 (51.90)	589 (34.36)	1259 (41.90)
5-6	223 (17.27)	327 (19.08)	550 (18.30)
7-8	151 (11.70)	288 (16.80)	439 (14.61)
>8	247 (19.13)	510 (29.75)	757 (25.19)
Number of hypertension-related diagnoses			
0	544 (42.14)	638 (37.22)	1182 (39.33)
1	442 (34.24)	614 (35.82)	1056 (35.14)
2	223 (17.27)	312 (18.21)	535 (17.81)
>2	82 (6.35)	150 (8.75)	232 (7.72)
Type of hypertension			
Normal BP[a, b]	200 (19.12)	307 (23.15)	507 (21.37)
High-normal BP[c]	458 (43.79)	564(42.53)	1022(43.09)
Grade 1 hypertension[d]	346 (33.08)	418 (31.52)	764 (32.21)
Grade 2 hypertension[e]	42 (4.01)	37 (2.79)	79 (3.33)
Total	1291 (43.00)	1714 (57.00)	3005 (100.00)

[a]BP: blood pressure.

[b]Normal BP: systolic BP<130 and diastolic BP<85 mmHg.

[c]High-normal BP: 130≤systolic BP≤139 and/or 85≤diastolic BP≤89 mmHg.

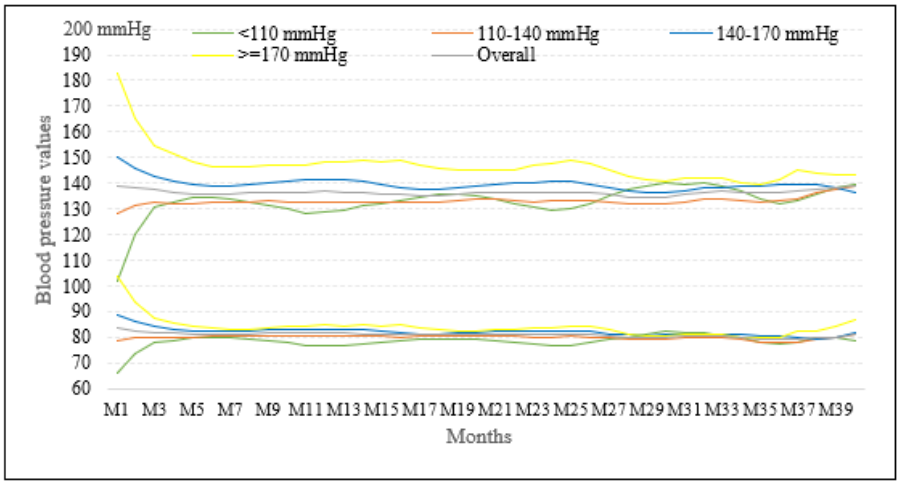
[d]Grade 1 hypertension: 140≤systolic BP≤159 and/or 90≤diastolic BP≤99 mmHg.

[e]Grade 2 hypertension: systolic BP≥160 and/or diastolic BP≥100 mmHg.

Trajectories of Monthly Mean BP Among Varied Cohorts

Figure 1 and Multimedia Appendix 2 demonstrate the changes in monthly mean SBP/DBP after different time periods (months) of HBPT among all 5658 participants and for patients with variable mean SBP/DBP in the first month. Both clusters of lines representing mean SBP/DBP featured a decreasing and "converging" trend, starting with a large gap between the highest and lowest mean SBP/DBP at the beginning and becoming closer and closer along the X-axis of months after the start of HBPT. The lines of mean SBP converged around a line just below 140 mmHg and the mean DBP line, just above 80 mmHg. The cohort with the highest start-month mean SBP (170+ mmHg) witnessed the greatest decrease in both SBP (from 183 mmHg in month 1 to 140 mmHg in month 35) and DBP (from 106 mmHg in month 1 to 79 mmHg in month 35). Conversely, the cohort with the lowest start-month mean SBP (110– mmHg) manifested the greatest increase in SBP (from 102 mmHg in month 1 to 141 mmHg in month 30) and DBP (from 66 mmHg in month 1 to 81 mmHg in month 40). The mean SBP/DBP among the cohort with the middle start-month mean SBP varied the least. The fastest decrease or increase occurred in the first 5-6 months.

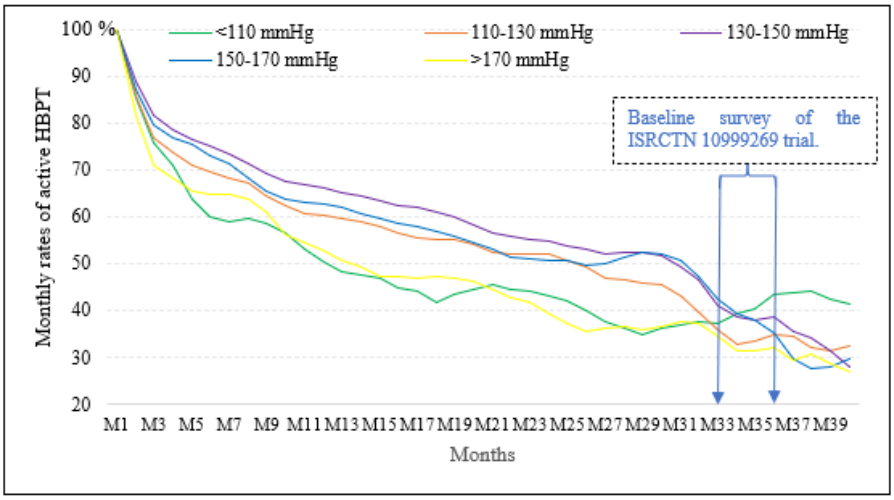
Figure 1. Monthly mean SBP/DBP among cohorts with varied start-month mean SBP. DBP: diastolic blood pressure; M1 through to M40: month 1 through to month 40; SBP: systolic blood pressure.



Rates of Monthly Active HBPT by Varied Start-Month BP

Figure 2 presents the rates of monthly active HBPT along the time axis in months. All 5658 participants performed HBPT in the first month, but then the rates dropped quickly for the next 2 to 4 months. The rates continued to decrease at a slower and slower pace subsequently. The patients with a start-month mean SBP of 130-150 mmHg displayed the highest rate of monthly active HBPT, followed by the 150-170 mmHg and 110-130 mmHg groups. The two extreme cohorts (the 110– and 170+ mmHg groups) were the least active in terms of HBPT. When patients were grouped according to their start-month mean DBP, the trajectories of monthly active HBPT rates mimicked the results shown in Figure 2 with respect to almost all features, except for narrower gaps between different groups (Multimedia Appendix 3).

Figure 2. Monthly rate of active HBPT by cohorts with varied start-month systolic blood pressure. HBPT: home blood pressure telemonitoring; M1 through to M40: month 1 through to month 40.



Multivariable Regression Modeling of SBP and DBP

Table 2 summarizes the statistics of our multivariable linear and percentile regression models for mean SBP and DBP in the last year. The linear regression analysis unveiled marginal and negative relations between the times of HBPT in the last year to both SBP (B=–0.09, *P*<.001) and DBP (B=–0.11, *P*<.001). In the percentile regression models, times of HBPT were found to have decreasing correlation coefficients for the two BP variables, from 0.16 to –0.35 and from 0.11 to –0.35 for SBP and DBP, respectively. In the percentile modeling, age also showed significant associations with SBP/DBP for all percentiles (positive for SBP and negative for DBP), whereas almost no significant relations were found for education, family history, and number of hypertension-related symptoms and diagnoses to both SBP and DBP (all *P*>.05). Sex was associated with DBP but not to SBP, whereas the duration of hypertension and BMI exhibited statistically significant links on apparently more percentiles for SBP than for DBP.

Table 2. Multivariable linear and percentile regression modeling of mean systolic blood pressure (SBP) and diastolic blood pressure (DBP).

Variables	All patients	Percentiles of mean SBP/DBP (%)								
		10	20	30	40	50	60	70	80	90
Systolic blood pressure										
(constant)										
Correlation coefficient	—[a]	–1.14	–0.75	–0.47	–0.26	–0.04	0.20	0.46	0.73	1.21
P value	.60	<.001	<.001	<.001	<.001	.15	<.001	<.001	<.001	<.001
Age										
Correlation coefficient	0.20	0.19	0.20	0.18	0.17	0.16	0.17	0.19	0.21	0.24
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Sex										
Correlation coefficient	–0.05	–0.05	–0.02	–0.03	–0.06	–0.06	–0.05	–0.04	–0.07	–0.08
P value	.07	.21	.61	.27	.04	.06	.08	.25	.06	.14
Education										
Correlation coefficient	–0.03	0.00	–0.01	–0.04	–0.05	–0.05	–0.06	–0.05	–0.05	0.00
P value	.24	.94	.76	.17	.06	.07	.04	.10	.16	.99
BMI										
Correlation coefficient	0.08	0.14	0.10	0.11	0.10	0.08	0.09	0.04	0.05	0.05
P value	.001	<.001	.001	<.001	<.001	.002	<.001	.11	.12	.33
Duration of hypertension										
Correlation coefficient	0.10	0.12	0.10	0.07	0.08	0.08	0.09	0.10	0.10	0.11
P value	<.001	<.001	.001	.007	.002	.001	.001	<.001	.003	.03
Family history of hypertension										
Correlation coefficient	0.01	0.06	–0.01	–0.02	–0.02	–0.03	–0.02	0.00	0.00	0.01
P value	.65	.08	.72	.49	.49	.23	.47	.90	.94	.83
Number of hypertension-related symptoms										
Correlation coefficient	–0.01	0.01	–0.06	–0.06	–0.03	–0.01	0.01	0.00	0.00	0.03
P value	.60	.78	.06	.02	.24	.80	.74	.94	.99	.59
Number of hypertension-related diagnoses										
Correlation coefficient	–0.02	0.01	0.00	0.01	0.01	0.00	–0.03	–0.04	–0.05	–0.08
P value	.50	.67	.89	.57	.81	.98	.22	.18	.13	.11
Annual measurement times										
Correlation coefficient	–0.09	0.16	0.10	0.04	0.01	–0.04	–0.11	–0.18	–0.27	–0.35
P value	<.001	<.001	<.001	.10	.64	.08	<.001	<.001	<.001	<.001
Diastolic blood pressure										
(constant)										
Correlation coefficient	—	–1.16	–0.74	–0.48	–0.22	–0.03	0.21	0.42	0.74	1.13
P value	.89	<.001	<.001	<.001	<.001	.18	<.001	<.001	<.001	<.001
Age										
Correlation coefficient	–0.23	–0.25	–0.23	–0.21	–0.20	–0.21	–0.24	–0.23	–0.23	–0.23
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Sex										
Correlation coefficient	–0.11	–0.10	–0.10	–0.06	–0.09	–0.07	–0.09	–0.13	–0.18	–0.16
P value	<.001	.02	.001	.04	.002	.02	.003	<.001	<.001	.002

Variables	All patients	Percentiles of mean SBP/DBP (%)								
		10	20	30	40	50	60	70	80	90
Education										
Correlation coefficient	–0.02	–0.03	–0.01	0.01	0.02	0.01	0.00	–0.02	–0.06	–0.05
P value	.54	.48	.66	.79	.51	.69	.95	.57	.08	.33
BMI										
Correlation coefficient	0.04	0.08	0.06	0.06	0.03	0.05	0.02	0.06	0.03	0.04
P value	.08	.02	.03	.04	.20	.03	.37	.02	.26	.39
Duration of hypertension										
Correlation coefficient	0.03	–0.04	0.04	0.04	0.03	0.05	0.04	0.03	0.05	0.02
P value	.17	.32	.14	.12	.20	.04	.12	.19	.13	.63
Family history of hypertension										
Correlation coefficient	0.02	0.05	0.02	0.03	0.02	0.00	0.00	–0.04	–0.02	0.01
P value	.49	.18	.46	.25	.45	.89	.90	.17	.52	.80
Number of hypertension-related symptoms										
Correlation coefficient	0.00	0.05	0.01	–0.03	–0.01	–0.03	–0.01	–0.01	0.01	0.01
P value	.92	.18	.71	.31	.70	.30	.63	.81	.84	.82
Number of hypertension-related diagnoses										
Correlation coefficient	–0.02	0.01	0.00	–0.02	–0.03	–0.04	–0.02	–0.01	–0.04	–0.01
P value	.39	.70	.92	.49	.21	.09	.38	.60	.21	.81
Annual measurement times										
Correlation coefficient	–0.11	0.11	0.01	–0.04	–0.04	–0.05	–0.13	–0.17	–0.24	–0.35
P value	<.001	.003	.78	.16	.10	.06	<.001	<.001	<.001	<.001

[a]Not applicable.

Multivariable Logistic Regression Models of BP Control Rate

Table 3 provides statistics of nine multivariable logistic regression models of BP control rate in the last year using different cut-off values (CVs) in dividing hypertensive patients into controlled (y=1 if a patient's BP control rate was greater than the CV) and uncontrolled (y=0 otherwise) categories. In terms of trend, times of HBPT displayed a consistent decreasing trend with BP control rate (correlation coefficient and odds ratio decreased from 0.53 and 1.7 in Model 1 to –0.62 and 0.54 in Model 9, respectively), whereas duration of hypertension presented a general increasing trend with the BP control rate from Models 1 to 9. The association of BP control rate was significant in the extreme models (Models 1, 2, 3, 4, 8, and 9) for times of HBPT, and was significant in the bottom models for age (Models 1 to 4) and number of hypertension-related diagnoses (Models 1 to 2), in top models for sex (Models 4 to 9) and BMI (from Models 7 to 9), in middle models for education (Models 3 to 5), and in all models for the duration of hypertension.

Table 3. Multivariable logistic regression modeling of blood pressure control rate.

Variables[a]	Model 1 (CV[b]=10%)	Model 2 (CV=20%)	Model 3 (CV=30%)	Model 4 (CV=40%)	Model 5 (CV=50%)	Model 6 (CV=60%)	Model 7 (CV=70%)	Model 8 (CV=80%)	Model 9 (CV=90%)
(constant)									
B[c]	1.25	0.61	0.27	–0.05	–0.35	–0.76	–1.24	–1.63	–2.49
OR[d]	3.50	1.84	1.30	0.95	0.71	0.47	0.29	0.20	0.08
P value	<.001	<.001	<.001	.30	<.001	<.001	<.001	<.001	<.001
Age									
B	0.17	0.19	0.19	0.12	0.08	0.03	–0.04	–0.05	–0.06
OR	1.19	1.20	1.21	1.13	1.09	1.03	0.96	0.95	0.94
P value	.007	.001	<.001	.02	.13	.55	.50	.49	.49
Sex									
B	0.12	0.09	0.09	0.14	0.11	0.16	0.19	0.18	0.20
OR	1.13	1.09	1.10	1.15	1.12	1.17	1.21	1.19	1.22
P value	.08	.14	.10	.01	.05	.01	.005	.02	.04
Education									
B	0.11	0.07	0.11	0.12	0.11	0.09	0.12	0.14	0.12
OR	1.11	1.07	1.12	1.13	1.12	1.09	1.13	1.15	1.13
P value	.12	.25	.05	.03	.05	.14	.08	.06	.22
BMI									
B	0.04	0.00	-0.01	–0.04	–0.06	–0.09	–0.14	–0.14	–0.17
OR	1.04	1.00	0.99	0.96	0.94	0.92	0.87	0.87	0.84
P value	.50	.94	.85	.37	.21	.10	.02	.03	.05
Duration of hypertension									
B	–0.19	–0.13	–0.11	–0.12	–0.16	–0.21	–0.25	–0.21	–0.31
OR	0.83	0.87	0.89	0.88	0.86	0.81	0.78	0.81	0.73
P value	.001	.007	.02	.01	.002	<.001	<.001	.004	.003
Family history of hypertension									
B	0.02	0.02	0.01	-0.02	0.05	0.08	0.03	0.00	–0.04
OR	1.02	1.02	1.01	0.98	1.05	1.09	1.03	1.00	0.96
P value	.67	.74	.81	.70	.29	.11	.58	.97	.61
Number of hypertension-related symptoms									
B	0.03	–0.02	–0.01	–0.02	–0.02	–0.01	–0.02	0.01	–0.08
OR	1.03	0.98	0.99	0.98	0.98	0.99	0.98	1.01	0.92
P value	.61	.68	.86	.65	.69	.89	.69	.92	.38
Number of hypertension-related diagnoses									
B	0.15	0.10	0.09	0.10	0.02	–0.03	–0.01	–0.06	0.04
OR	1.16	1.11	1.09	1.10	1.02	0.97	0.99	0.94	1.04
P value	.01	.04	.08	.05	.63	.51	.88	.36	.64
Annual measurement times									
B	0.53	0.24	0.15	0.12	0.03	0.02	–0.09	–0.22	–0.62
OR	1.70	1.27	1.16	1.12	1.03	1.02	0.92	0.80	0.54
P value	<.001	<.001	.002	.02	.58	.75	.12	<.001	<.001

[a]The dependent variable in Models 1 to 9 was assigned 1 if the blood control rate of the patient under concern was greater than the CV or 0 otherwise.

[b]CV: cut-off value of blood pressure control rate.

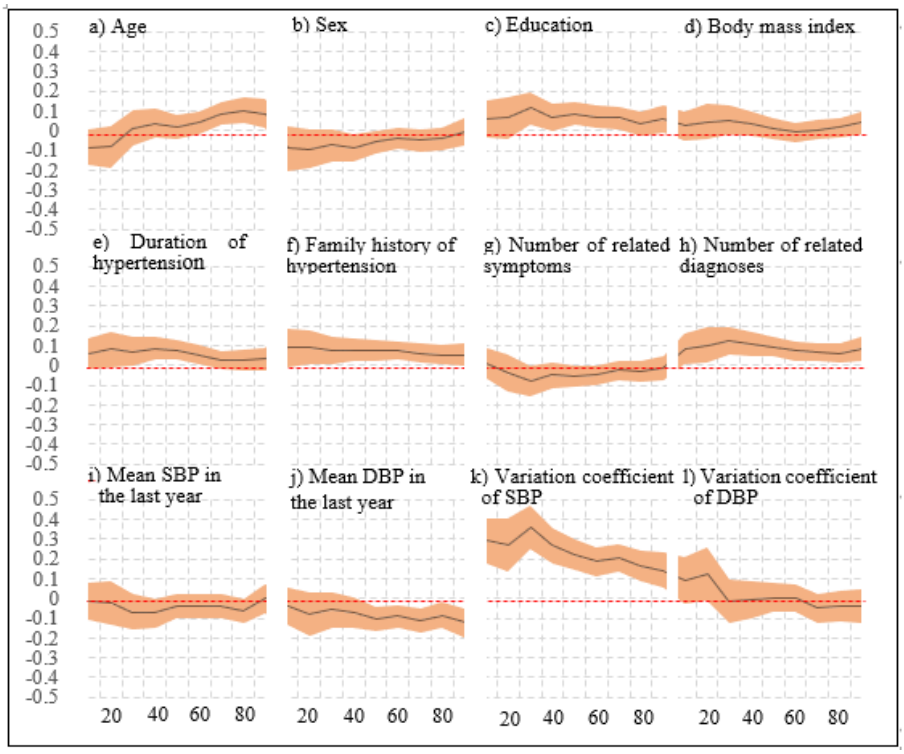
[c]B: correlation coefficient.

[d]OR: odds ratio.

Multivariable Percentile Regression Analysis of HBPT

Figure 3 displays, in shaded curves, the multivariable percentile regression coefficients between times of HBPT in the last year and the independent variables studied. Of all the curves, only those representing the variation coefficients of SBP (Figure 3k) and number of hypertension-related diagnoses (Figure 3h) presented a clear distance from the dashed red line (B=0) along all of the percentiles, and only the curve representing the mean SBP overlapped with the red line along the entire percentile axis (Figure 3i). All of the remaining shaded curves demonstrated interceptions with the red line for a larger or smaller part of the percentiles. The variation coefficients of SBP (Figure 3k) manifested the largest distance from the red line and exhibited a decreasing trend along the percentiles, whereas the coefficient of age presented an increasing trend. The main statistics of the multivariable percentile regression model are given in Multimedia Appendix 4.

Figure 3. Multivariable percentile regression modeling of factors affecting times of home blood pressure telemonitoring (HBPT). The y-axis represents the regression coefficient. The x-axis represents quantiles of times of HBPT in the last year. DBP: diastolic blood pressure; SBP: systolic blood pressure.



Discussion

Effects of HBPT on SBP/DBP

Our study unveiled novel and meaningful BP trajectories after HBPT among hypertensive cohorts with varied mean SBP in the first month (Figure 1). Instead of simply lowering SBP or DBP as documented in most previous related studies, HBPT followed mixed changes in our study depending on the resultant B*P* values of the patients under concern. The magnitude of changes ranged from –43 to +39 mmHg for SBP and from –27 to +15 mmHg for DBP. When controlled for commonly reported confounder variables such as sex, age, education, duration of hypertension, BMI, family history of hypertension, and numbers of hypertension-related symptoms and diagnoses, the differentiated effects of HBPT on SBP/DBP were still observable. In the multivariable percentile regression model (Table 2), times of HBPT showed moderate to strong relations with both SBP and DBP. In our multivariable logistic regression models (Table 3), times of HBPT again demonstrated strong and differentiated associations with BP control rate. These findings suggest that HBPT played major yet bidirectional or "target-converging" roles. Interestingly, the "target" here was the widely validated and accepted defining values of hypertension control (ie, SBP below 140 and/or DBP under 90 mmHg [11,20]). When the monitored SBP/DBP was higher than the target, HBPT may have urged the patients to take actions to reduce their BP through self-titrating antihypertensive medication; consulting their doctors for initiating or intensifying antihypertensive treatment; and practicing more rigorous lifestyle changes that have been shown to reduce hypertension, including weight loss, dietary approaches, and physical activity [21-24]. For patients with lower than the target BP, HBPT may have informed them to consult their doctors for milder treatment agents or doses, or to perhaps reduce further self-management efforts.

Implications of the "target-converging" effect remain to be carefully examined. It is well-established that "convergence" downward from above the "target" is beneficial to patients via various mechanisms, including a lower risk of cerebral hemorrhage [25,26]. An upward "convergence" from far below the "target" (eg, SBP<110 mmHg) may also result in better health outcomes due to, for instance, reduced chances of cerebral ischemia [27]. However, upward "convergence" from a certain range below the "target" (eg, SBP from 130 mmHg to 140 mmHg) may be harmful to patients' health.

Determinants of HBPT

The declining and varied rates of monthly active HBPT for different cohorts (Figure 2) suggest that SBP/DBP and time may be the most important factors affecting HBPT. The reasons why the middle cohort (subgroups with a start-month mean SBP=130-150 mmHg in Figure 2 or DBP=80-90 mmHg in Multimedia Appendix 3) exhibited the highest rates of monthly HBPT may be because their resultant SBP/DBP levels were the closest to the defining values of hypertension control. The closer a patient's BP is to the defining value, the higher their chances to obtain meaningful feedback (success or failure in hypertension management) from an HBPT, and thus the greater the desire to perform the monitoring. Conversely, participants in the two extreme cohorts may have become either frustrated with or relieved to perform HBPT.

The decreasing trend over time following start of the HBPT project may be mainly attributed to increasing familiarity with the resultant SBP/DBP. In other words, when the patients' ability to anticipate the results enhanced, their desire or interest in performing HBPT decreased. This is consistent with our findings (Figure 3 and Multimedia Appendix 4) that the variation coefficients of SBP were independently linked to the times of HBPT.

Our multivariable percentile regression model also identified independent associations between HBPT and age, education, duration of hypertension, family history, and diagnosis of hypertension complications. Perceived risk may be the main reason underlying these relations. In other words, patients of older age, with better education, a longer duration of hypertension, more diagnoses, and family history may perceive themselves at an elevated risk for developing hypertension complications and thus become more active in HBPT [28,29]. It is worth noting that all of these correlations were weaker than those of the values of DBP and variations in SBP in terms of the magnitude of the correlation coefficient or the duration of percentiles.

Variations in Relationships

Our study uncovered interesting variations in the relationships between HBPT and its influencing factors. Times of HBPT presented negative associations with mean DBP (Figure 3j), but did not show statistically significant associations with mean SBP (Figure 3i). This may be explained by a dynamic interaction between the dependent and independent variables. More specifically, more frequent HBPT led to greater chances for identifying elevated BP, which in turn led to greater efforts to reduce BP and then to greater decreases in BP, and finally to nonsignificant relations between SBP and HBPT. The same dynamics could also be in play for DBP but led to negative associations, since a substantial portion of the patients had isolated systolic hypertension and their DBP was indirectly reduced via the interactions between HBPT and SBP. The variation in SBP (Figure 3k) could not be as easily reduced as SBP/DBP via the interaction dynamics, and thus showed consistent strong correlations with HBPT. The nonsignificant relations between variation in DBP and HBPT (Figure 3l) may be related to the much smaller value as compared with the variation in SBP, which thus attracted relatively little attention from the patients.

Similar variations in correlations were also observed in the models using SBP/DBP as the dependent variables. For example, age showed a sustained and positive association with SBP but a continuous negative link to DBP. These contradictory relations have been reported in various hypertensive populations, especially those dominated by relatively older patients with isolated systolic hypertension [30,31]. In addition, sex was associated with DBP but not SBP, while BMI and duration of hypertension showed stronger links with SBP than with DBP. These findings are also similar with those of previous studies [32,33]. With regard to the BP control rate, our logistic regression model suggested that age, sex, and education were protective factors; BMI and number of hypertension-related diagnoses were risk factors; and the effects of these factors were complex, being observable in various parts of the models. The mechanisms and implications of these phenomena merit further exploration.

Strengths and Limitations

Our study has both strengths and limitations. This study used data from a relatively large-scale (5658 patients with hypertension) and long-term cohort. Relatively in-depth analysis of the determinants of HBPT was performed, with particular attention paid to subtle and differential interactions with the resultant BP outcome. This study thus produced useful trajectories of monthly mean SBP/DBP and monthly active rates of HBPT for up to 40 months. Multivariable linear, percentile, and logistic regression modeling of times of HBPT, mean SBP/DBP, and BP control rate as the dependent variables, respectively, enabled cross-checks and comparisons of the results.

This study also suffers from drawbacks. First, being performed at home by ordinary residents, the B*P* values from HBPT are prone to various influences. Second, the study population was relatively old (65.50 years on average) and the findings should be generalized with caution. Third, the study considered only SBP/DBP as the outcome variables without considering others (eg, complications, health care burden) and lacked comparison with patients who had not used HBPT. Fourth, BP readings are susceptible to diurnal and intraobserver variations, which can lead to measurement biases, although our use of monthly and hourly mean SBP/DBP may have helped to reduce these biases to some extent. Our further research activities in response to these shortcomings include performing household surveys/observations to help identify the factors influencing HBPT readings, extend the HBPT to younger populations, and

perform further analyses linking HBPT with major adverse cardiovascular events (eg, apoplexy) and quality of life.

Conclusions

HBPT had major and "target-converging" effects on SBP/DBP. The "target" was the widely validated and accepted defining values of hypertension control (ie, SBP below 140 and/or DBP under 90 mmHg). HBPT was followed by SBP/DBP reductions or increases for cohorts with a mean BP higher or lower than the "target," respectively. The magnitude of changes was a few times greater than commonly documented. These differentiated effects remained observable into the third year after initiation of HBPT. BP, variation in BP, and time were the most important determinants of HBPT uptake, whereas age, education, duration of hypertension, family history, and diagnosis of hypertension complications were also linked to the uptake but at apparently weaker strength. HBPT displayed stronger associations with the variation in SBP than in DBP.

There is a clear need for differentiated thinking over the application and assessment of HBPT. First, the traditional approach of simply comparing the effects in the intervention group as a whole with that in the control group is prone to underestimation of the actual influences of HBPT, since decreases in a portion of the patients were offset by increases in others. HBPT leads to BP decrease, stability, or increase depending on the complex and dynamic context of the patient under concern. These varied effects may not necessarily all be beneficial and merit careful scrutiny in the future. This study thus highlights the need for correcting outdated beliefs or practices and leveraging new opportunities with the application of HBPT. Second, the difference in the "white coat" effect suggests lower than traditional cut-off values of hypertension control when readings from HBPT were used. In other words, patients should be better educated about the "white coat" effect and that they need to exert further efforts to maintain their HBPT readings slightly below 140/90 mmHg. Third, the varied responses toward different levels of HBPT readings indicate selective telemonitoring, group-specific "targets," or even personalized interventions. Fourth, relatively less attention paid to DBP than to SBP implies that additional efforts are needed to promote balanced awareness among patients. In particular, patients should be informed that DBP is as important as SBP and thus merits equal attention in self-monitoring.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (grant 72004002). The funding source has not played any role in the study design, analysis, or in the decision to submit the manuscript for publication.

Authors' Contributions

QX and XZ contributed equally in conceiving this study and drafting this manuscript. RL, XG, and GL implemented the computational analysis. LZ and QW facilitated project implementation. DW accessed and verified all the data in the study. XS provided expertise for design of the study and revised and finalized the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Related questions used in the baseline household survey and value assignment.
[DOCX File , 16 KB - jmir_v24i10e37648_app1.docx]

Multimedia Appendix 2
Monthly mean SBP/DBP among cohorts with varied start-month mean DBP. DBP: diastolic blood pressure; SBP: systolic blood pressure; M1 through to M40: month 1 through to month 40.
[PNG File , 16 KB - jmir_v24i10e37648_app2.png]

Multimedia Appendix 3
Monthly rate of active HBPT by cohorts with varied start-month diastolic blood pressure. HBPT: home blood pressure telemonitoring; M1 through to M40: month 1 through to month 40.
[PNG File , 22 KB - jmir_v24i10e37648_app3.png]

Multimedia Appendix 4
Multivariable linear and percentile regression coefficients of times of home blood pressure telemonitoring (HBPT) .
[DOCX File , 23 KB - jmir_v24i10e37648_app4.docx]

References

1. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
2. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 Practice Guidelines for the management of arterial hypertension of the European Society of Hypertension and the European Society of Cardiology: ESH/ESC Task Force for the Management of Arterial Hypertension. J Hypertens 2018 Dec;36(12):2284-2309. [doi: 10.1097/HJH.0000000000001961] [Medline: 30379783]
3. Villar R, Sánchez RA, Boggia J, Peñaherrera E, Lopez J, Barroso WS, et al. Recommendations for home blood pressure monitoring in Latin American countries: A Latin American Society of Hypertension position paper. J Clin Hypertens 2020 Apr;22(4):544-554. [doi: 10.1111/jch.13815] [Medline: 32049425]
4. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-67; discussion 467. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]
5. Omboni S, Guarda A. Impact of home blood pressure telemonitoring and blood pressure control: a meta-analysis of randomized controlled studies. Am J Hypertens 2011 Sep;24(9):989-998. [doi: 10.1038/ajh.2011.100] [Medline: 21654858]
6. Duan Y, Xie Z, Dong F, Wu Z, Lin Z, Sun N, et al. Effectiveness of home blood pressure telemonitoring: a systematic review and meta-analysis of randomised controlled studies. J Hum Hypertens 2017 Jul;31(7):427-437. [doi: 10.1038/jhh.2016.99] [Medline: 28332506]
7. Omboni S, Ferrari R. The role of telemedicine in hypertension management: focus on blood pressure telemonitoring. Curr Hypertens Rep 2015 Apr;17(4):535. [doi: 10.1007/s11906-015-0535-3] [Medline: 25790799]
8. Shimbo D, Artinian NT, Basile JN, Krakoff LR, Margolis KL, Rakotz MK, American Heart Associationthe American Medical Association. Self-measured blood pressure monitoring at home: a Joint Policy Statement from the American Heart Association and American Medical Association. Circulation 2020 Jul 28;142(4):e42-e63. [doi: 10.1161/CIR.0000000000000803] [Medline: 32567342]
9. McManus RJ, Mant J, Haque MS, Bray EP, Bryan S, Greenfield SM, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. JAMA 2014 Aug 27;312(8):799-808. [doi: 10.1001/jama.2014.10057] [Medline: 25157723]
10. Zullig LL, Melnyk SD, Goldstein K, Shaw RJ, Bosworth HB. The role of home blood pressure telemonitoring in managing hypertensive populations. Curr Hypertens Rep 2013 Aug;15(4):346-355 [FREE Full text] [doi: 10.1007/s11906-013-0351-6] [Medline: 23625207]
11. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 International Society of Hypertension global hypertension practice guidelines. J Hypertens 2020 Jun;38(6):982-1004. [doi: 10.1097/HJH.0000000000002453] [Medline: 32371787]
12. 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]
13. Yi SS, Tabaei BP, Angell SY, Rapin A, Buck MD, Pagano WG, et al. Self-blood pressure monitoring in an urban, ethnically diverse population: a randomized clinical trial utilizing the electronic health record. Circ Cardiovasc Qual Outcomes 2015 Mar;8(2):138-145 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.114.000950] [Medline: 25737487]
14. 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]
15. Lu X, Yang H, Xia X, Lu X, Lin J, Liu F, et al. Interactive mobile health intervention and blood pressure management in adults. Hypertension 2019 Sep;74(3):697-704. [doi: 10.1161/HYPERTENSIONAHA.119.13273] [Medline: 31327259]
16. Omboni S, Panzeri E, Campolo L. E-Health in hypertension management: an insight into the current and future role of blood pressure telemonitoring. Curr Hypertens Rep 2020 Jun 06;22(6):42. [doi: 10.1007/s11906-020-01056-y] [Medline: 32506273]
17. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
18. Verberk WJ, Kessels AGH, Thien T. Telecare is a valuable tool for hypertension management, a systematic review and meta-analysis. Blood Press Monit 2011 Jun;16(3):149-155. [doi: 10.1097/MBP.0b013e328346e092] [Medline: 21527847]
19. Shen X, Xiao S, Liu R, Tong G, Liu T, Wang D. Personalized hypertension management based on serial assessment and telemedicine (PHMA): a cluster randomize controlled trial protocol in Anhui, China. BMC Cardiovasc Disord 2021 Mar 12;21(1):135 [FREE Full text] [doi: 10.1186/s12872-021-01943-5] [Medline: 33711941]
20. Cuspidi C, Tadic M, Grassi G, Mancia G. Treatment of hypertension: The ESH/ESC guidelines recommendations. Pharmacol Res 2018 Feb;128:315-321. [doi: 10.1016/j.phrs.2017.10.003] [Medline: 29080798]

21. Agarwal R, Bills JE, Hecht TJW, Light RP. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. Hypertension 2011 Jan;57(1):29-38. [doi: 10.1161/HYPERTENSIONAHA.110.160911] [Medline: 21115879]
22. Omboni S. Connected health in hypertension management. Front Cardiovasc Med 2019;6:76. [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
23. Mancia G, Parati G. Home blood pressure monitoring: a tool for better hypertension control. Hypertension 2011 Jan;57(1):21-23. [doi: 10.1161/HYPERTENSIONAHA.110.163188] [Medline: 21115877]
24. Fletcher BR, Hartmann-Boyce J, Hinton L, McManus RJ. The effect of self-monitoring of blood pressure on medication adherence and lifestyle factors: a systematic review and meta-analysis. Am J Hypertens 2015 Oct;28(10):1209-1221. [doi: 10.1093/ajh/hpv008] [Medline: 25725092]
25. Pancorbo O, Rodriguez-Luna D. Blood pressure lowering in acute intracerebral hemorrhage. Aging 2018 Nov 07;10(11):3056-3057 [FREE Full text] [doi: 10.18632/aging.101637] [Medline: 30404950]
26. Anderson CS, Selim MH, Molina CA, Qureshi AI. Intensive blood pressure lowering in intracerebral hemorrhage. Stroke 2017 Jul;48(7):2034-2037 [FREE Full text] [doi: 10.1161/STROKEAHA.117.016185] [Medline: 28626061]
27. Lattanzi S, Silvestrini M. Reader response: Cerebral ischemia and deterioration with lower blood pressure target in intracerebral hemorrhage. Neurology 2019 Apr 16;92(16):776. [doi: 10.1212/WNL.0000000000007310] [Medline: 30988091]
28. Flynn SJ, Ameling JM, Hill-Briggs F, Wolff JL, Bone LR, Levine DM, et al. Facilitators and barriers to hypertension self-management in urban African Americans: perspectives of patients and family members. Patient Prefer Adherence 2013;7:741-749. [doi: 10.2147/PPA.S46517] [Medline: 23966772]
29. Konlan KD, Afam-Adjei CJ, Afam-Adjei C, Oware J, Appiah TA, Konlan KD, et al. Practice and sociodemographic factors influencing self-monitoring of blood pressure in Ghanaians with hypertension. Int J Chronic Dis 2020;2020:6016581. [doi: 10.1155/2020/6016581] [Medline: 32566645]
30. Franklin SS. Systolic blood pressure: it's time to take control. Am J Hypertens 2004 Dec;17(12 Pt 2):49S-54S. [doi: 10.1016/j.amjhyper.2004.08.020] [Medline: 15607435]
31. Satoh M, Metoki H, Asayama K, Murakami T, Inoue R, Tsubota-Utsugi M, et al. Age-related trends in home blood pressure, home pulse rate, and day-to-day blood pressure and pulse rate variability based on longitudinal cohort data: The Ohasama Study. J Am Heart Assoc 2019 Aug 06;8(15):e012121 [FREE Full text] [doi: 10.1161/JAHA.119.012121] [Medline: 31333055]
32. Hosseini M, Baikpour M, Yousefifard M, Fayaz M, Koohpayehzadeh J, Ghelichkhani P, et al. Blood pressure percentiles by age and body mass index for adults. EXCLI J 2015;14:465-477 [FREE Full text] [doi: 10.17179/excli2014-635] [Medline: 26417366]
33. Bernabe-Ortiz A, Carrillo-Larco RM, Miranda JJ. Association between body mass index and blood pressure levels across socio-demographic groups and geographical settings: analysis of pooled data in Peru. PeerJ 2021;9:e11307. [doi: 10.7717/peerj.11307] [Medline: 33976985]

Abbreviations

BP: blood pressure
CV: cut-off value
DBP: diastolic blood pressure
HBPT: home blood pressure telemonitoring
RCT: randomized controlled trial
SBP: systolic blood pressure

Edited by R Kukafka; submitted 02.03.22; peer-reviewed by R Armstrong Junior, G Wu; comments to author 23.08.22; revised version received 29.08.22; accepted 16.09.22; published 11.10.22.

Please cite as:
Xue Q, Zhang X, Liu R, Guan X, Li G, Zhao L, Wang Q, Wang D, Shen X
Differentiated Effects and Determinants of Home Blood Pressure Telemonitoring: Three-Year Cohort Study in Jieshou, Anhui, China
J Med Internet Res 2022;24(10):e37648
URL: https://www.jmir.org/2022/10/e37648
doi:10.2196/37648
PMID:36114000

Original Paper

The Preferences of Transgender and Nonbinary People for Virtual Health Care After the COVID-19 Pandemic in Canada: Cross-sectional Study

Jose M Navarro[1], BSc; Ayden I Scheim[1,2], PhD; Greta R Bauer[1], MPH, PhD

[1]Department of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

[2]Department of Epidemiology and Biostatistics, Dornsife School of Public Health, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Jose M Navarro, BSc
Department of Epidemiology and Biostatistics
Schulich School of Medicine & Dentistry
Western University
Western Centre for Public Health and Family Medicine, 3rd Fl
1465 Richmond Street
London, ON, N6G 2M1
Canada
Phone: 1 647 678 5673
Email: jnavarr9@uwo.ca

Abstract

Background: Virtual health care use has dramatically increased in response to the COVID-19 pandemic, raising the question of its potential role after the pandemic. For transgender (trans) and nonbinary (TNB) people, virtual care is promising because it may expand access to appropriate health care providers. However, emerging research indicates potential disparities in virtual care access related to sociodemographic, health, and social factors. There is a paucity of research on the factors affecting patient preferences for virtual versus in-person care, particularly in TNB communities.

Objective: This study aimed to identify the sociodemographic, health, and social factors associated with postpandemic virtual care preferences in TNB communities.

Methods: The 2020 Trans PULSE Canada COVID survey examined the health, social, and economic impacts of the COVID-19 pandemic among 820 TNB participants who previously completed the prepandemic 2019 Trans PULSE Canada survey (n=2783). Data were weighted to the demographics of the 2019 sample. Chi-square tests were used to compare postpandemic preferences for virtual versus in-person care across sociodemographic, health, and social characteristics. Participants provided open-text responses explaining their preferences, which were used to contextualize quantitative findings.

Results: Among 812 participants who indicated whether they would prefer virtual or in-person care after the pandemic, a weighted 32.7% (n=275) would prefer virtual care and 67.3% (n=537) would prefer in-person care. Preference for in-person over virtual care was associated with being in the 14-19 (49/56, weighted 85.0%), 50-64 (51/62, weighted 80.0%), and ≥65 (9/10, weighted 90.7%) age groups (χ^2_5=19.0; *P*=.002). Preference for virtual over in-person care was associated with having a chronic health condition (125/317, weighted 37.7% versus 150/495, weighted 29.9%; χ^2_1=4.7; *P*=.03) and having probable anxiety (229/645, weighted 34.7% versus 46/167, weighted 25.7%; χ^2_1=4.3; *P*=.04). Among participants with romantic partners, preferences varied based on the partner's level of support for gender identity or expression (χ^2_3=13.3; *P*=.004). Participants with moderately supportive partners were more likely than participants with very supportive partners to prefer in-person care (36/43, weighted 85.1% versus 275/445, weighted 62.3%). Care preferences did not vary significantly based on the indicators of socioeconomic status. Open-text responses showed that multiple factors often interacted to influence participant preferences, and that some factors, such as having a chronic condition, simultaneously led some participants to prefer virtual care and others to prefer in-person care.

Conclusions: TNB people may have differential interest in virtual care based on factors including age, chronic and mental health conditions, and gender-unsupportive home environments. Future research examining virtual care preferences would benefit from mixed methods intersectional approaches across these factors, to explore complexity in the barriers and facilitators of virtual

care access and quality. These observed differences support flexibility with options to choose between in-person and virtual health care to meet TNB patients' specific health needs.

(J Med Internet Res 2022;24(10):e40989) doi:10.2196/40989

KEYWORDS

virtual care; telemedicine; telehealth; eHealth; transgender; gender identity; COVID-19; gender-affirming care; older adult; mental health; chronic condition; social support

Introduction

Virtual care, also known as telemedicine or telehealth, has historically been underused due to concerns about quality of care, insufficient regulatory frameworks, lack of technological infrastructure, and its limited inclusion in public health insurance programs in countries with publicly funded health systems like Canada [1-3]. In 2020, the COVID-19 pandemic forced health care institutions to adapt services for remote delivery and drove Canadian provincial and territorial governments to rapidly strengthen public health insurance coverage for virtual care [4-6].

Moving forward from the pandemic, transgender (trans) and nonbinary (TNB) communities may benefit from increased access to virtual care [7]. People who are trans have a gender identity that differs from the sex they were assigned at birth [8,9]. People who are nonbinary, who may or may not identify as trans, have gender identities beyond woman or man [9]. TNB people often struggle to find health care providers who are clinically and culturally competent to address TNB health issues [8], and clinics that specialize in gender-affirming care (eg, hormone therapy) tend to be limited to major urban centers [7,9]. TNB people also frequently report experiences of stigma and discrimination while traveling to and accessing health care, leading to care avoidance and unmet health care needs [8,10]. Virtual care could help address these concerns by allowing TNB people to access a broader array of both general and gender-affirming health services and practitioners, regardless of their place of residence [1,7,11]. Travel costs and lost income would also be reduced by virtual care delivery [1,7,11,12]. Current gender-affirming virtual care options like Connect-Clinic [13] in Ontario show promise in this regard, but they are still limited, with demand far exceeding supply [7]. Moreover, gender-affirming care is only one subset of the health care needed by TNB patients.

Despite its promise, the adoption of virtual care must be approached with caution as current research on its benefits and drawbacks is relatively limited and provides mixed results [1,3,6,14]. Access to virtual care may be particularly limited for TNB people who lack a reliable internet connection, lack safe confidential spaces where they can access virtual care, or lack the digital literacy needed to effectively use online health services [11,12,14-17]. Although public funding of virtual care has expanded in Canada, it is still limited [6]. For instance, at the time of writing, most jurisdiction codes only cover care delivered via video call and not via secure text messaging [3], and some provinces place a cap on the number of virtual appointments that physicians can bill to public health insurance [6]. Along with parallel for-profit models of telemedicine, this limited coverage could translate into socioeconomic disparities in virtual care access for TNB people [2-4,15].

With the goal of informing postpandemic virtual care practice and policies, this paper draws on data from a national survey of TNB people in Canada conducted during the COVID-19 pandemic to examine how preferences indicated for postpandemic in-person versus virtual care varied based on sociodemographic, health-related, and social characteristics. We also drew on qualitative data from open-text fields to elucidate the reasons for participants' preferences.

Methods

Participants and Procedures

Administered from September 21 to October 20, 2020, the Trans PULSE Canada COVID survey collected national data on the health, economic, and social impacts of the COVID-19 pandemic on TNB people in Canada [18]. Participants were recruited from a list of 1187 TNB Canadian residents aged ≥14 years (as of 2019) who consented for recontact after completing the 2019 Trans PULSE Canada survey. Details of the methods for the 2019 study have been published previously. In brief, participants were recruited via convenience sampling and completed the survey on paper, by telephone, or online [19]. Eligible participants for the 2020 COVID survey were contacted via their preferred communication method among email, telephone, text, or letter mail and directed to a webpage with further explanation of the study. Of the 1187 people contacted, 820 (69.1%) completed the 2020 survey. Consent was implied by survey completion. All questionnaires were self-administered online in English or French through REDCap [20], although participants were also offered the option of a mailed questionnaire and of receiving accessibility supports such as translation. Survey questions were pretested for clarity, and participants could skip any question they did not wish to answer. Participants provided separate (optional) consent for the publication of quotes from open-text fields. A CAD $20 (US $15) gift card honorarium was offered to each participant who completed the 2020 COVID survey.

Ethics Approval

The Research Ethics Boards at Western University, Drexel University, and Wilfrid Laurier University approved the 2020 Trans PULSE Canada COVID survey (Western University: project ID 116072; Drexel University: protocol number 2005007801; Wilfrid Laurier University: REB number 6557).

Measures

Primary Outcome

Virtual care was defined for participants as "health care or medical advice delivered via phone call, video call, or text message." Preference for virtual versus in-person health care was indicated by the response to the following question: "In general, would you prefer virtual over in-person care when COVID-19 is no longer an issue?" Those who responded "yes" were categorized as preferring virtual care and those who responded "no" were categorized as preferring in-person care. Participants were also asked to explain the reasons for their preference in an open-response text box.

Sociodemographic, Health-Related, and Social Factors

Sociodemographic characteristics included age, gender, racialization, indigeneity, province or territory, rurality, and indicators of socioeconomic status (annual income, employment status, low-income household, and housing stability). Racialized participants were those who self-identified or were perceived or treated as people of color in Canada. Those living in a town or municipality with a population less than 10,000 were categorized as rural, based on postal code. The definition of a low-income household was based on Statistics Canada's low-income measure [21], and examples of unstable housing included living in a shelter, motel, or car.

Health-related factors included self-identification as disabled, chronic conditions, virtual care experiences, mental health conditions, and gender-affirming care status. Participants were identified as having a chronic condition if they indicated that they had chronic pain, a chronic illness, or a chronic health condition. Participants were asked the following 4 questions related to virtual care experiences: (1) whether they had accessed virtual care since March 12, 2020, the start of the COVID-19 pandemic; (2) the type of virtual care received (physical, mental, or other including gender-affirming care); (3) the platform over which they received virtual care (phone call, video call, texting, or other including email); and (4) whether they had avoided virtual care due to their TNB identity since the start of the COVID-19 pandemic. Current mental health was assessed using validated scales. Anxiety symptoms were measured using the Overall Anxiety Severity and Impairment Scale (OASIS) [22]. Participants responded to 5 items, each with 5 options (coded 0-4 with a possible score range of 0-20), indicating the relative frequency or intensity of their anxiety symptoms in the past week. Summed scores of 8 or above indicated probable anxiety [22]. Depressive symptoms were measured with the 10-item abridged Center for Epidemiologic Studies Depression Scale (CES-D-10) [23]. Each item had 4 response options (coded 0-3 with a possible score range of 0-30), with higher summed scores reflecting a greater frequency of depressive symptomatology. A cutoff score of 10 or above indicated potential clinical depression [23]. Participants were also asked the extent to which they had received or were considering gender-affirming medical care, which was defined in the survey as including "puberty blockers, gender-affirming hormones, surgeries, or body modifications."

Participants responded to questions assessing their social environment and relationships, that is, whether they had experienced intimate partner violence since August 2019, how supportive their partner(s) and parent(s)/guardian(s) were of their gender identity or expression, and whether they were concerned about family stress from confinement and violence at home during the pandemic.

Analysis

As less than 1% of COVID survey participants (8 of 820) did not respond to the question on postpandemic preferences for virtual care, these participants were excluded from the analysis. Responses to the 2020 COVID survey were weighted to match the demographic profile of the full 2019 sample on characteristics like age, ethnoracial background, and socioeconomic status, using a raking algorithm. Weights were used in case loss to follow-up between the 2019 and COVID surveys was nonrandom and to allow for better comparability between the pre-COVID and COVID samples. Rao-Scott chi-square tests with α=.05 were performed on weighted data to compare preferences for virtual versus in-person care after the pandemic across sociodemographic, health, and social factors. All quantitative analyses were carried out in SAS 9.4. software (SAS Institute Inc).

Similar to a sequential explanatory design [24], direct quotations from participants' open-text responses are included in the results, with indications of the cited participant's age, gender, and province to contextualize quantitative findings. Themes corresponding to the statistically significant quantitative results were identified. Open-text responses explaining virtual care preferences were sorted into these themes.

Results

Virtual Care Preferences and Identified Themes

Of 812 participants, a weighted 32.7% (n=275) said they would prefer virtual care after the COVID-19 pandemic, while 67.3% (n=537) would prefer in-person care. Most participants (746/812, 91.9%) provided an explanation for their virtual care preference in the open-response question. Based on the quantitative findings that follow, the following broad themes were identified, into which the open-text responses were sorted: age, disability and chronic conditions, mental health, social environment, and the logistics of care access (eg, convenience and technological literacy). After further examination of open-text responses, we identified discrimination and stigma as another prominent theme not present in the quantitative results. Table 1 presents the unweighted sociodemographic characteristics of the 812 participants included in the analysis.

Table 1. Sociodemographic characteristics of the analytic sample.

Characteristic	Total sample (N=812), n (%)[a]
Age (years)	
14-19	56 (6.9)
20-24	147 (18.2)
25-34	306 (38.0)
35-49	225 (27.9)
50-64	62 (7.7)
≥65	10 (1.2)
Gender	
Woman or girl	200 (24.7)
Man or boy	198 (24.4)
Indigenous or cultural gender identity	18 (2.2)
Nonbinary or similar	395 (48.7)
Racialization	
Yes	108 (13.3)
No	702 (86.7)
Indigenous in Canada	
Yes	59 (7.3)
No	750 (92.7)
Immigration status	
Newcomer (past 5 years)	23 (2.8)
Immigrant (nonnewcomer)	91 (11.2)
Born in Canada	698 (86.0)
Province of residence	
Alberta	151 (18.6)
Atlantic[b]	38 (4.7)
British Columbia	176 (21.7)
Manitoba	27 (3.3)
Newfoundland and Labrador	6 (0.7)
Ontario	305 (37.6)
Quebec	81 (10.0)
Saskatchewan	26 (3.2)
Territories[c]	1 (0.1)
Rural	
Yes	47 (5.8)
No	762 (94.2)
Personal annual income (CAD$[d]; age≥16 years)	
None	142 (17.8)
<$14,999	214 (26.8)
$15,000-$29,999	160 (20.1)
$30,000-$49,999	124 (15.5)
$50,000-$79,999	94 (11.8)

Characteristic	Total sample (N=812), n (%)[a]
≥$80,000	64 (8.0)
Education	
Less than high school	44 (5.4)
High school diploma	62 (7.6)
Some college or university	206 (25.4)
College or university degree	354 (43.6)
Graduate/professional degree	145 (17.9)
Employment situation (age ≥25 years)	
Permanent full-time	220 (37.0)
Employed, not permanent full-time	206 (34.6)
Not employed or on leave	136 (22.9)
Not employed and student or retired	33 (5.5)
Low-income household(past year; age≥16 years)	
Yes	310 (41.1)
No	444 (58.9)
Housing stability	
Stable	806 (99.3)
Unstable	6 (0.7)
Disability identity	
Yes	217 (26.7)
No	595 (73.3)

[a]Unweighted frequencies and proportions are reported.

[b]Including New Brunswick, Nova Scotia, and Prince Edward Island.

[c]Including Northwest Territories, Nunavut, and Yukon.

[d]A currency exchange rate of CAD $1=US $0.75 is applicable.

Sociodemographic Differences

Age was associated with care preferences (χ^2_5=19.0; *P*=.002), with a larger proportion of participants aged 14-19 years (49/56, weighted 85.0%), 50-64 years (51/62, weighted 80.0%), and ≥65 years (9/10, weighted 90.7%) preferring in-person care after the pandemic compared with participants in other age groups (Table 2). Gender identity was also associated with care preference (χ^2_3=11.2; *P*=.01), with participants self-identifying with an Indigenous or culturally specific gender minority identity being more likely to prefer virtual care than those identifying as women, men, or nonbinary, although no significant difference in virtual care preference was observed by indigenous identity (χ^2_1=0.02; *P*=.90). No significant associations were identified between care preferences and the various indicators of socioeconomic status.

Table 2. Preference for virtual care versus in-person care by sociodemographic characteristics in Trans PULSE Canada COVID survey participants.

Characteristic	Prefers virtual care (N=275)		Prefers in-person care (N=537)		*P* value[a]
	Value, n (%)[b]	95% CI[b]	Value, n (%)[b]	95% CI[b]	
Age (years)					.002
14-19	7 (15.0)	4.6-25.4	49 (85.0)	74.6-95.4	
20-24	52 (34.1)	25.9-42.2	95 (65.9)	57.8-74.1	
25-34	114 (37.1)	31.3-42.9	192 (62.9)	57.1-68.7	
35-49	90 (38.7)	31.7-45.6	135 (61.3)	54.4-68.3	
50-64	11 (20.0)	8.4-31.5	51 (80.0)	68.5-91.6	
≥65	1 (9.3)	0.0-26.9	9 (90.7)	73.1-100.0	
Gender					.01
Woman or girl	61 (29.1)	22.4-35.9	139 (70.9)	64.1-77.6	
Man or boy	56 (26.1)	19.6-32.6	142 (73.9)	67.4-80.4	
Indigenous or cultural gender identity	10 (55.4)	31.3-79.5	8 (44.6)	20.5-68.7	
Nonbinary or similar	148 (37.0)	31.9-42.2	247 (63.0)	57.8-68.1	
Racialization					.69
Yes	38 (34.5)	24.9-44.2	70 (65.5)	55.8-75.1	
No	236 (32.4)	28.7-36.2	466 (67.6)	63.8-71.3	
Indigenous in Canada					.90
Yes	22 (33.6)	21.0-46.3	37 (66.4)	53.7-79.0	
No	253 (32.8)	29.2-36.4	497 (67.2)	63.6-70.8	
Province of residence					__[c]
Alberta	49 (31.4)	23.5-39.2	102 (68.6)	60.8-76.5	
Atlantic[d]	9 (24.1)	9.9-38.2	29 (75.9)	61.8-90.1	
British Columbia	70 (40.1)	32.3-47.8	106 (59.9)	52.2-67.7	
Manitoba	9 (27.1)	10.3-43.8	18 (72.9)	56.2-89.7	
Newfoundland and Labrador	4 (73.7)	38.9-100.0	2 (26.3)	0.0-61.1	
Ontario	96 (31.0)	25.4-36.5	209 (69.0)	63.5-74.6	
Quebec	28 (30.6)	20.0-41.1	53 (69.4)	58.9-80.0	
Saskatchewan	9 (31.3)	12.3-50.2	17 (68.8)	49.8-87.7	
Territories[e]	1 (100.0)	100.0-100.0	0 (0.0)	0.0-0.0	
Rural					.12
Yes	20 (43.8)	28.5-59.1	27 (56.2)	40.9-71.5	
No	254 (32.1)	28.5-35.6	508 (67.9)	64.4-71.5	
Personal annual income (CAD$[f]; age≥16 years)					.08
None	50 (34.9)	26.3-43.4	92 (65.1)	56.6-73.7	
<$14,999	78 (34.0)	27.2-40.7	136 (66.0)	59.3-72.8	
$15,000-$29,999	46 (26.9)	19.7-34.2	114 (73.1)	65.8-80.3	
; $30,000-$49,999	42 (34.7)	25.6-43.8	82 (65.3)	56.2-74.4	
$50,000-$79,999	42 (45.6)	34.9-56.4	52 (54.4)	43.6-65.1	
≥$80,000	16 (24.2)	12.9-35.5	48 (75.8)	64.5-87.1	
Employment situation (age ≥25 years)					.40
Permanent full-time	87 (39.3)	32.4-46.3	133 (60.7)	53.7-67.6	

Characteristic	Prefers virtual care (N=275)		Prefers in-person care (N=537)		*P* value[a]
	Value, n (%)[b]	95% CI[b]	Value, n (%)[b]	95% CI[b]	
Employed, not permanent full-time	69 (32.9)	26.0-39.9	137 (67.1)	60.1-74.0	
Not employed or on leave	48 (34.5)	25.7-43.3	88 (65.5)	56.7-74.3	
Not employed and student or retired	9 (25.4)	9.2-41.6	24 (74.6)	58.4-90.8	
Low-income household(past year; age≥16 years)					.23
Yes	114 (35.5)	29.8-41.3	196 (64.5)	58.7-70.2	
No	145 (31.1)	26.5-35.7	299 (68.9)	64.3-73.5	
Housing stability					.46
Stable	273 (32.6)	29.1-36.1	533 (67.4)	63.9-70.9	
Unstable	2 (48.1)	4.3-92.0	4 (51.9)	8.0-95.7	

[a]Comparing care preferences across sociodemographic characteristics using the Rao-Scott chi-square test.

[b]Proportions are weighted to the sociodemographics of the prepandemic Trans PULSE Canada sample.

[c]*P* value was not available owing to small cell sizes.

[d]Including New Brunswick, Nova Scotia, and Prince Edward Island.

[e]Including Northwest Territories, Nunavut, and Yukon.

[f]A currency exchange rate of CAD $1=US $0.75 is applicable.

Disability and Chronic Conditions

Preference for virtual care was more common in participants with chronic conditions (125/317, weighted 37.7% versus 150/495, weighted 29.9%; χ^2_1=4.7; *P*=.03) (Table 3). Although no significant differences in care preferences were found based on self-identification as disabled (χ^2_1=0.43; *P*=.51), 1 participant who preferred virtual over in-person care noted:

> *With COVID, things are going virtual and it is so helpful. I am more connected now, and people make more of an effort for virtual visits as well. This is the kind of access people living with my disability need.* [Nonbinary or similar, aged 25-34 years, living in Ontario]

At the same time, others cited their chronic health conditions as a reason to prefer in-person care:

> *I think virtual visits would be great long term for things like prescription refills or ordering things that don't require a physical inspection. But with chronic health concerns I feel being visually seen in person for a checkup throughout the year is vital to my staying healthy and functional.* [Man, aged 25-34 years, living in Ontario]

These sentiments echo another common theme in participant open-text responses—that preferences for virtual versus in-person care were contingent on the specific type of care being sought. Numerous participants noted a preference for virtual care for appointments they deemed as not requiring in-person treatment (like prescription refills), and a preference for in-person care when they felt that aspects like physical examination were necessary.

Open-text responses also showed that the examined sociodemographic, health-related, and social factors often interacted to affect virtual care preferences. For instance, 1 participant discussed how their chronic condition exacerbated geographic barriers to in-person care:

> *Taking public transit to and from an appointment when I'm already not feeling well can take a substantial amount of energy. Having to do this over and over again for ongoing investigation into my symptoms takes even more. I desperately need this energy for basic caretaking of myself… I shouldn't have to make myself sicker to access health care.* [Nonbinary or similar, aged 25-34 years, living in Ontario]

Table 3. Disability and chronic conditions, virtual care experiences, mental health factors, and social environment of Trans PULSE Canada COVID cohort participants preferring virtual versus in-person care.

Characteristic	Prefers virtual care (N=275)		Prefers in-person care (N=537)		*P* value[a]
	Value, n (%)[b]	95% CI[b]	Value, n (%)[b]	95% CI[b]	
Disability identity					.51
Yes	78 (34.8)	28.0-41.5	139 (65.2)	58.5-72.0	
No	197 (32.1)	28.1-36.2	398 (67.9)	63.8-71.9	
Chronic conditions				.	.03
Yes	125 (37.7)	32.0-43.5	192 (62.3)	56.5-68.0	
No	150 (29.9)	25.5-34.2	345 (70.1)	65.8-74.5	
Virtual care access since March 12, 2020					<.001
Yes	200 (38.4)	33.8-43.0	299 (61.6)	57.0-66.2	
No	75 (24.5)	19.3-29.6	238 (75.5)	70.4-80.7	
Type of virtual care[c]					
Physical health care	173 (39.3)	34.3-44.4	250 (60.7)	55.6-65.7	.42
Mental health care	118 (39.6)	33.4-45.8	164 (60.4)	54.2-66.6	.56
Other	5 (57.8)	25.4-90.2	5 (42.2)	9.8-74.6	.22
Virtual care platform[c]					
Phone call	176 (38.0)	33.1-42.9	264 (62.0)	57.1-66.9	.67
Video call	113 (42.0)	35.4-48.6	139 (58.0)	51.4-64.6	.14
Texting/SMS text messaging	19 (32.7)	18.9-46.5	36 (67.3)	53.5-81.1	.41
Other	1 (100.0)	100.0-100.0	0 (0.0)	0.0-0.0	__[d]
Virtual care avoidance because of trans/nonbinary identity since March 12, 2020					.14
Yes	48 (38.9)	29.7-48.2	76 (61.1)	51.8-70.3	
No	227 (31.6)	27.9-35.4	461 (68.4)	64.6-72.1	
Anxiety					.04
Probable anxiety (OASIS[e] ≥8)	229 (34.7)	30.7-38.7	416 (65.3)	61.3-69.3	
No probable anxiety (OASIS <8)	46 (25.7)	18.6-32.8	121 (74.3)	67.2-81.4	
Depression					.06
Probable depression (CES-D-10[f] ≥10)	242 (34.2)	30.3-38.0	445 (65.8)	62.0-69.7	
No probable depression (CES-D-10 <10)	33 (25.1)	17.0-33.2	92 (74.9)	66.8-83.0	
Gender-affirming medical care status					.52
Had all needed care	84 (28.6)	22.8-34.4	198 (71.4)	65.6-77.2	
In the process of completing	101 (33.6)	27.9-39.4	187 (66.4)	60.6-72.1	
Planning, but not begun	32 (38.4)	27.4-49.3	54 (61.6)	50.7-72.6	
Unsure if going to seek care	27 (33.2)	22.0-44.4	46 (66.8)	55.6-78.0	
Not planning	31 (35.4)	24.7-46.1	52 (64.6)	53.9-75.3	
Experienced intimate partner violence (since August 2019)					
Yes	38 (26.2)	18.5-33.9	93 (73.8)	66.1-81.5	.10
No	234 (33.8)	30.0-37.7	439 (66.2)	62.3-70.0	
Spouse/partner support of gender identity or expression[g]					.004
Very supportive	170 (37.7)	32.8-42.6	275 (62.3)	57.4-67.2	
Not very or somewhat supportive	7 (14.9)	4.3-25.6	36 (85.1)	74.4-95.7	

Characteristic	Prefers virtual care (N=275)		Prefers in-person care (N=537)		*P* value[a]
	Value, n (%)[b]	95% CI[b]	Value, n (%)[b]	95% CI[b]	
Not at all supportive	5 (69.1)	36.2-100.0	3 (30.9)	0.0-63.8	
Does not know about gender identity/expression	1 (20.0)	0.0-58.8	2 (80.0)	41.2-100.0	
Parent/guardian support of gender identity or expression[g]					.06
Very supportive	78 (29.0)	23.0-35.1	177 (71.0)	64.9-77.0	
Not very or somewhat supportive	122 (34.0)	28.7-39.3	227 (66.0)	60.7-71.3	
Not at all supportive	35 (47.2)	35.0-59.4	40 (52.8)	40.6-65.0	
Does not know about gender identity/expression	27 (31.9)	20.9-42.8	54 (68.1)	57.2-79.1	
Concerned about family stress from confinement due to COVID-19					.09
Extremely or very	111 (35.4)	29.6-41.2	194 (64.6)	58.8-70.4	
Somewhat	74 (27.2)	21.4-33.1	192 (72.8)	66.9-78.6	
Not at all	89 (35.5)	29.0-42.0	150 (64.5)	58.0-71.0	
Concerned about violence at home during COVID-19					.98
Extremely or very	9 (31.9)	13.1-50.7	16 (68.1)	49.3-86.9	
Somewhat	18 (31.6)	18.1-45.1	36 (68.4)	54.9-81.9	
Not at all	248 (32.9)	29.2-36.5	485 (67.1)	63.5-70.8	

[a]Comparing care preferences across experiences with virtual care, mental health factors, and social environment factors using the Rao-Scott chi-square test.

[b]Proportions are weighted to the sociodemographics of the prepandemic Trans PULSE Canada sample.

[c]Among those who received virtual care since March 12, 2020 (n=499).

[d]*P* value was not available owing to small cell sizes.

[e]OASIS: Overall Anxiety Severity and Impairment Scale.

[f]CES-D-10: 10-item abridged Center for Epidemiologic Studies Depression Scale.

[g]Results reported only for participants who indicated that these questions were applicable to them.

Virtual Care Experiences

Among participants who accessed virtual care since the pandemic (n=499, weighted 59.4%), 38.4% (weighted, 200/499) indicated a preference for virtual care after the pandemic, a significantly greater proportion than participants who did not access virtual care (75/313, weighted 24.5%; χ^2_1=14.4; *P*<.001) (Table 3). Among those who accessed virtual care, postpandemic preferences did not vary depending on whether they accessed it for physical (χ^2_1=0.66; *P*=0.42) or mental (χ^2_1=0.35; *P*=0.56) health care, or whether they received care via phone (χ^2_1=0.18; *P*=0.67), video call (χ^2_1=2.2; *P*=0.14), or texting (χ^2_1=0.68; *P*=0.41). Multiple participants mentioned that they received gender-affirming care, specifically hormone therapy, via virtual means, and that email was another platform through which they received care, which was not listed in our survey. Postpandemic virtual care preferences did not vary based on whether participants had avoided virtual care during the pandemic due to their TNB identities (χ^2_1=2.2; *P*=.14).

Mental Health Factors

A majority (n=645, weighted 78.3%) of the sample had probable anxiety (indicated by OASIS scores≥8), and participants with probable anxiety were more likely to prefer virtual care after the pandemic than participants without (229/645, weighted 34.7% versus 46/167, weighted 25.7%; χ^2_1=4.3; *P*=.04) (Table 3). Similarly, 84.3% (weighted, n=687) of the sample had CES-D-10 scores≥10, indicating clinically significant depressive symptomatology. A higher proportion of participants reaching this cutoff preferred virtual care, although this difference only approached statistical significance (242/687, weighted 34.2% versus 33/125, weighted 25.1%; χ^2_1=3.5; *P*=.06).

Consistent with our quantitative results, mental health conditions acted as barriers to accessing in-person treatment for some participants, with 1 participant saying:

> *It is sometimes difficult to leave the house or go to new environments without assistance given [my] anxiety and mental health issues - having the option to do virtual [care] makes health care more accessible.* [Nonbinary, aged 20-24 years, living in British Columbia]

At the same time, mental health concerns had different implications for care among other participants. One participant stated the following:

> *Phone conversations make me anxious, and I will avoid [them] just because it's going to be a phone conversation. Video calls are worse than phone calls.*

> *I feel like a lot of physical health issues need to be seen or felt, and mental health issues are better conveyed in person.* [Nonbinary, aged 25-34 years, living in British Columbia]

While not captured in our quantitative results, several participants also cited gender dysphoria (discomfort arising when one's physical characteristics do not align with their gender identity [6]) as a reason for preferring in-person care. Some participants explained that they experienced gender dysphoria while attending appointments via video call because they had to see themselves on screen. With regard to telephone appointments, 1 participant expressed the following:

> *My voice is higher than I would like, so talking over the phone makes me dysphoric about how the professional on the other line sees me based on my voice.* [Nonbinary, aged 20-24 years, living in Saskatchewan]

Social Environment

Most TNB participants with a spouse or romantic partner had one who was very supportive of their gender identity or expression (445/499, weighted 89.0%) (Table 3). Spouse or partner support was associated with care preference (χ^2_3=13.3; *P*=.004), with participants having moderately supportive partners being more likely to prefer in-person care than participants having very supportive partners (36/43, weighted 85.1% versus 275/445, weighted 62.3%). No significant differences depending on having experienced intimate partner violence were found (χ^2_1=2.7; *P*=.10).

Participants with very supportive parents or guardians more frequently indicated a preference for in-person care (177/255, weighted 71.0%) compared with other participants, although there were no statistically significant differences (χ^2_3=7.5; *P*=.06). However, multiple participants, particularly youth, mentioned that a lack of parental support was a reason to prefer in-person care. One participant made the following statement:

> *I live with my transphobic parents and wouldn't feel comfortable having medical appointments with them nearby.* [Nonbinary, aged 14-19 years, living in Ontario]

Privacy issues with other family members and roommates were also cited in open-text responses as reasons to avoid virtual care.

Participants preferring virtual care also attributed their preference to previously experienced discrimination in health care settings. One participant explained:

> *As a trans person, healthcare settings are a place where I've experienced a lot of abuse and oppression. My providers now are mostly good, but the setting is triggering… I don't miss not having to expose myself to that.* [Woman, aged 35-49 years, living in British Columbia]

In contrast, numerous participants stated that they were more likely to be misgendered (referred to as the wrong gender [9]) in virtual care settings, justifying a preference for in-person care. One participant made the following statement:

> *To be honest, it has been tough being misgendered constantly in my home over virtual meetings. I'd prefer in-person care elsewhere, so home becomes a safer space again, with less misgendering.* [Nonbinary or similar, aged 25-34 years, living in British Columbia]

Other participants explicitly noted that their TNB identity was irrelevant to their virtual care preferences.

Discussion

This paper identifies factors that may influence postpandemic preferences for virtual versus in-person care among TNB people in Canada. While most participants preferred in-person care, around 1 in 3 (weighted 32.7%) indicated a postpandemic preference for virtual care. Lack of access to virtual care during the pandemic was associated with postpandemic preference for in-person care, highlighting the importance of identifying and addressing the challenges that certain populations disproportionately face while attempting to access telemedicine. Participants who were aged 14-19 or ≥50 years were more likely to prefer in-person care over virtual care compared with other age groups. Other research has shown a lower level of digital literacy among older adults as a factor contributing to preference for in-person care [11,15-17]. Given that all participants in this study completed the survey online and therefore likely had substantial digital literacy, our finding of a similar age-related preference suggests that additional factors may be at play. Consistent with previous findings [16], chronic conditions and anxiety symptoms were associated with preferring virtual care over in-person care, suggesting that telemedicine offers a promising alternative for those whose health conditions prevent them from safely and comfortably attending in-person appointments. Older adults, with a higher prevalence of chronic conditions compared with the general population [25], may therefore particularly stand to benefit from telemedicine.

Previous research predicted that unsupportive home environments could compromise the privacy and sense of safety necessary to access virtual care [11], and our results in general support this with regard to a lack of support for gender identity or expression in the home. Having gender-unsupportive romantic partners was associated with preference for in-person care. Associations with gender support from parents or guardians were less clear, which may be because relevance is age dependent. Privacy issues may in part explain why participants preferring virtual care were less likely to be adolescents (age 14-19 years), who may have no option but to live with family or roommates, including those who are unsupportive (or who do not know). However, we did not measure whether our participants were actually living with their romantic partners, or their parents or guardians at the time of participation. Regardless, these privacy concerns highlight how a move to virtual care could disadvantage those in unsafe or unsupportive home environments, especially younger people who may not have the freedom or financial means to live on their own.

It is important to note that our quantitative results only captured the overall preferences of the sample, and closer examination of qualitative results showed heterogeneity. For example, some

participants reported that anxiety prevented them from accessing in-person care, whereas others had more anxiety surrounding virtual care. While this heterogeneity was expected, the broad operationalizations used for some of our variables may have obscured salient differences. For instance, participants were categorized based on identity as disabled, which does not capture the wide diversity in disability experiences. Accordingly, some participants with physical disabilities cited barriers to accessing in-person care, whereas others reported that disabilities like autism made virtual care less accessible. Similar variability was found for mental health and other chronic conditions. Future studies would benefit from examining these conditions with more nuanced and detailed categorization, for instance, by distinguishing between the types of disabilities, chronic conditions, and mental health conditions.

Another limitation of this study was that, because the survey focused broadly on the COVID-19 experiences of TNB people, only a few questions assessed virtual care experiences. Thus, some key factors like the ability to find a private space to access virtual care were not directly measured. Further, the survey's online administration mode may have introduced selection bias favoring those with preferences for virtual care. In general, although this study used the largest national sample of TNB people in Canada during the COVID-19 pandemic, the results should be interpreted with caution and are not generalizable to all TNB communities given that participants were a subset of a convenience sample.

A strength of this study was its use of qualitative responses to elucidate and elaborate upon quantitative results. While open-ended survey items generally do not produce the rich data of a true qualitative study, these responses highlighted how the factors under study could not be discussed in isolation from one another. Rather, our limited quotes suggest that they act together, for example, the ways a chronic health condition intersects with geographic barriers to require repeated public transit trips that then exacerbate the health condition. This suggests that future research should draw on an intersectionality theoretical framework to explore these processes of interaction and how they may relate to social power [26]. Future qualitative research would contribute to a deeper understanding of the processes through which virtual care preferences and access play out across intersections of gender, race, disability, chronic disease, age, and socioeconomic status, and a mixed methods intersectional approach would aid in the development of more nuanced and well-rounded virtual care policies and programs.

Some participants expressed through open-text responses that their TNB identity was relevant to their virtual care preferences, mentioning gender dysphoria, transphobia in healthcare, or misgendering. Similar proportions of participants who did and did not avoid virtual care during the pandemic due to their TNB identity reported postpandemic preference for virtual care. This finding suggests that those who avoided virtual care due to their TNB identity, for instance, due to anticipated discrimination [8], may similarly avoid in-person care. However, many participants did not mention their gender, with some explicitly stating that their preference was unrelated to their TNB identity. Because many of our quantitative findings and open-text responses were related to non-TNB-specific factors, they may be applicable to the broader population.

The heterogeneity within our results suggests that flexibility in choice regarding modality of care delivery may best support the diverse needs of patients, particularly those in TNB communities. For example, because some participants justified preferences for in-person care based on anxiety, chronic conditions, and parental support, while others used these factors to justify virtual care preferences, practitioners should be prepared to deliver care via virtual or in-person means, if feasible. Flexibility should also be afforded in terms of the mode of virtual care delivery. Some TNB participants experienced dysphoria when seeing themselves on a video call, but others may have been more concerned with practitioners misgendering them based on how their voice sounded over telephone appointments.

Depending on the province or territory in Canada where physicians practice, they may face limitations on the types and modes of services they can bill to public health insurance, which may influence the care options that they are willing or able to provide [3,6]. In certain jurisdictions, physicians have daily caps on the number of virtual appointments they can bill to public health insurance [6]. While previous studies found that patients often want care delivered over secure text messaging, when possible, few jurisdictions in Canada offer publicly funded texting services [3]. Only certain types of care are publicly insured, with, for instance, walk-in appointments not being covered in Nova Scotia [6]. Additionally, many changes made in the Canadian virtual care policy in response to the COVID-19 pandemic are still temporary [6]. To improve care quality and access, policymakers and funders may consider implementing permanent funding schemes that cover a wider range of services and reward physicians similarly for the same services delivered via different modes.

Although virtual care may be valuable to address existing disparities in care access for TNB-specific and nonspecific services, both virtual and in-person care options will remain important after the pandemic. Researchers and care providers should continue to identify populations for whom virtual care is particularly beneficial, identify the barriers that may prevent them from accessing it, and propose and appraise interventions designed to overcome these disparities.

Acknowledgments

The Trans PULSE Canada Study Team would like to acknowledge and thank the trans and nonbinary people who generously shared their time and experience with us. The authors also thank Dr Hannah Kia (University of British Columbia) for her guidance regarding the qualitative component of this paper, and Dr Alisa Grigorovich (Brock University) for her assistance with the

conceptualization of this manuscript. The Trans PULSE Canada Study was funded by the Canadian Institutes of Health Research (funding reference number PJT-159690). The funder played no role in the conduct or interpretation of this research.

Authors' Contributions

AIS and GRB are the co-principal investigators of the Trans PULSE Canada Study and contributed to the survey design and data collection. JMN, AIS, and GRB performed the data analysis. JMN wrote the manuscript, and all co-authors contributed to its revision.

Conflicts of Interest

None declared.

References

1. Kahn JM. Virtual visits — Confronting the challenges of telemedicine. N Engl J Med 2015 Apr 30;372(18):1684-1685. [doi: 10.1056/nejmp1500533]
2. Owens B. Telemedicine on the rise but lagging in Canada. CMAJ 2018 Sep 24;190(38):E1149-E1150 [FREE Full text] [doi: 10.1503/cmaj.109-5634] [Medline: 30249766]
3. Buying Access Will Cost You: The unintended consequences of for-profit virtual care. The College of Family Physicians of Canada. URL: https://www.cfpc.ca/CFPC/media/Resources/Health-Policy/Corporatization-of-Care.pdf [accessed 2022-06-23]
4. Ortega G, Rodriguez JA, Maurer LR, Witt EE, Perez N, Reich A, et al. Telemedicine, COVID-19, and disparities: Policy implications. Health Policy Technol 2020 Sep;9(3):368-371 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.001] [Medline: 32837888]
5. Telemedicine and virtual care guidelines (and other clinical resources for COVID-19). Royal College of Physicians and Surgeons of Canada. URL: https://www.royalcollege.ca/rcsite/documents/about/covid-19-resources-telemedicine-virtual-care-e#rcr [accessed 2021-06-03]
6. Virtual care in Canada: progress and potential. Canadian Medical Association. URL: https://www.cma.ca/sites/default/files/2022-02/Virtual-Care-in-Canada-Progress-and-Potential-EN.pdf [accessed 2022-06-23]
7. Stoehr JR, Hamidian Jahromi A, Hunter EL, Schechter LS. Telemedicine for gender-affirming medical and surgical care: A systematic review and call-to-action. Transgender Health 2022 Apr 01;7(2):117-126. [doi: 10.1089/trgh.2020.0136]
8. Kachen A, Pharr JR. Health Care Access and Utilization by Transgender Populations: A United States Transgender Survey Study. Transgend Health 2020 Sep;5(3):141-148 [FREE Full text] [doi: 10.1089/trgh.2020.0017] [Medline: 33644308]
9. Puckett JA, Cleary P, Rossman K, Newcomb ME, Mustanski B. Barriers to Gender-Affirming Care for Transgender and Gender Nonconforming Individuals. Sex Res Social Policy 2018 Mar;15(1):48-59 [FREE Full text] [doi: 10.1007/s13178-017-0295-8] [Medline: 29527241]
10. Giblon R, Bauer GR. Health care availability, quality, and unmet need: a comparison of transgender and cisgender residents of Ontario, Canada. BMC Health Serv Res 2017 Apr 18;17(1):283 [FREE Full text] [doi: 10.1186/s12913-017-2226-z] [Medline: 28420361]
11. Hamnvik OR, Agarwal S, AhnAllen CG, Goldman AL, Reisner SL. Telemedicine and Inequities in Health Care Access: The Example of Transgender Health. Transgender Health 2022 Apr 01;7(2):113-116. [doi: 10.1089/trgh.2020.0122]
12. Nanda M, Sharma R. A Review of Patient Satisfaction and Experience with Telemedicine: A Virtual Solution During and Beyond COVID-19 Pandemic. Telemed J E Health 2021 Dec;27(12):1325-1331. [doi: 10.1089/tmj.2020.0570] [Medline: 33719577]
13. Connect Clinic. URL: https://www.connect-clinic.com [accessed 2022-06-26]
14. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
15. Agarwal P, Wang R, Meaney C, Walji S, Damji A, Gill N, et al. Sociodemographic differences in patient experience with primary care during COVID-19: results from a cross-sectional survey in Ontario, Canada. BMJ Open 2022 May 09;12(5):e056868 [FREE Full text] [doi: 10.1136/bmjopen-2021-056868] [Medline: 35534055]
16. Choi NG, DiNitto DM, Marti CN, Choi BY. Telehealth Use Among Older Adults During COVID-19: Associations With Sociodemographic and Health Characteristics, Technology Device Ownership, and Technology Learning. J Appl Gerontol 2022 Mar;41(3):600-609 [FREE Full text] [doi: 10.1177/07334648211047347] [Medline: 34608821]
17. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient Characteristics Associated With Choosing a Telemedicine Visit vs Office Visit With the Same Primary Care Clinicians. JAMA Netw Open 2020 Jun 01;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]
18. COVID Cohort T1 2020 Survey – English – Paper Version. Trans PULSE Canada. URL: https://transpulsecanada.ca/wp-content/uploads/2021/08/COVID-cohort-survey.pdf [accessed 2022-01-13]

19. Scheim AI, Coleman T, Lachowsky N, Bauer GR. Health care access among transgender and nonbinary people in Canada, 2019: a cross-sectional survey. CMAJ Open 2021;9(4):E1213-E1222 [FREE Full text] [doi: 10.9778/cmajo.20210061] [Medline: 34933879]
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
21. Table 4.2 Low-income measures thresholds (LIM-AT and LIM-BT) for private households of Canada, 2015. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t4_2-eng.cfm [accessed 2022-01-13]
22. Campbell-Sills L, Norman SB, Craske MG, Sullivan G, Lang AJ, Chavira DA, et al. Validation of a brief measure of anxiety-related severity and impairment: the Overall Anxiety Severity and Impairment Scale (OASIS). J Affect Disord 2009 Jan;112(1-3):92-101 [FREE Full text] [doi: 10.1016/j.jad.2008.03.014] [Medline: 18486238]
23. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D. American Journal of Preventive Medicine 1994 Mar;10(2):77-84. [doi: 10.1016/s0749-3797(18)30622-6]
24. Ivankova NV, Creswell JW, Stick SL. Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. Field Methods 2016 Jul 21;18(1):3-20. [doi: 10.1177/1525822X05282260]
25. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. . [doi: 10.1016/S0140-6736(16)31678-6]
26. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. Am J Public Health 2012 Jul;102(7):1267-1273. [doi: 10.2105/AJPH.2012.300750] [Medline: 22594719]

Abbreviations

CES-D-10: 10-item abridged Center for Epidemiologic Studies Depression Scale
OASIS: Overall Anxiety Severity and Impairment Scale
TNB: transgender (trans) and nonbinary

Edited by A Mavragani; submitted 23.07.22; peer-reviewed by S Hagens, OP Hamnvik; comments to author 11.08.22; revised version received 08.09.22; accepted 09.09.22; published 26.10.22.

Please cite as:
Navarro JM, Scheim AI, Bauer GR
The Preferences of Transgender and Nonbinary People for Virtual Health Care After the COVID-19 Pandemic in Canada: Cross-sectional Study
J Med Internet Res 2022;24(10):e40989
URL: https://www.jmir.org/2022/10/e40989
doi:10.2196/40989
PMID:36170497

Original Paper

Digital Technology Access and Health-Related Internet Use Among People Experiencing Homelessness in Hungary: Quantitative Survey

Nóra Radó[1], MA; Edmond Girasek[1], PhD; Sándor Békási[2,3], MSc, MD; Zsuzsa Győrffy[1], PhD

[1]Institute of Behavioural Sciences, Faculty of Medicine, Semmelweis University, Budapest, Hungary

[2]Health Centre, Hungarian Charity Service of the Order of Malta, Budapest, Hungary

[3]Telemedicine Workgroup, FitPuli Kft., Győr, Hungary

Corresponding Author:
Nóra Radó, MA
Institute of Behavioural Sciences
Faculty of Medicine
Semmelweis University
Nagyvárad sqr 4
Budapest, 1089
Hungary
Phone: 36 205679360
Email: rado.nora@gmail.com

Abstract

Background: In recent years, there has been an increase in the use of digital technology for personal health and well-being. Previous research has revealed that these technologies might provide vulnerable populations, including those who are homeless, better access to health services and thus a greater chance of more personalized care.

Objective: However, little is known about the relationship between technology and health among people experiencing homelessness in Central and Eastern Europe. This study is part of a series of studies by the Digital Health Research Group at Semmelweis University (Budapest, Hungary) in cooperation with the Hungarian Charity Service of the Order of Malta; it aims to assess the existing technological resources available for the homeless population and their health-related internet use characteristics to set the ground for potential health policy interventions, enabling better access to health services by strengthening the digital components of the existing health care system.

Methods: Between April 19, 2021, and August 11, 2021, a total of 662 people from 28 institutions providing social services for people experiencing homelessness in Budapest, Hungary, were surveyed about their access to digital tools and internet use patterns. For selected questions, the responses of a representative sample of the Hungarian population were used for comparison as the reference group. Chi-square tests and logistic regression analyses were performed to identify variables affecting internet use for health-related reasons.

Results: The results demonstrated a considerable level of internet use in the homeless population; 52.9% (350/662) of the respondents used the internet frequently compared with 81.3% (1220/1500) of the respondents in the reference group. Among the homeless group, 69.6% (461/662) of the respondents reported mobile phone ownership, and 39.9% (264/662) of the respondents added that it had a smartphone function. Moreover, 11.2% (70/662) of the respondents had already used a health mobile app, and 34.6% (229/662) of the respondents had used the internet for medical purposes. On the basis of these characteristics, we were able to identify a broadly defined, digitally engaged group among people experiencing homelessness (129/662, 19.5%). This subpopulation was inclined to benefit from digitalization related to their personal health. Multivariate analysis demonstrated that internet use for health reasons was more significant for younger respondents, women, those with higher levels of education, and those with no chronic conditions.

Conclusions: Although compared with the general population, health-related internet use statistics are lower, our results show that the idea of involving homeless populations in the digital health ecosystem is viable, especially if barriers to access are systematically reduced. The results show that digital health services have great promise as another tool in the hands of community shelters for keeping homeless populations well ingrained in the social infrastructure as well as for disease prevention purposes.

(J Med Internet Res 2022;24(10):e38729) doi:10.2196/38729

KEYWORDS

homelessness; digital technology; internet; access; internet use; homeless shelter; digital equity; mobile phone

Introduction

Homelessness in Hungary

Homelessness is a complex set of social, economic, and health challenges at both the individual and community levels. The term itself represents a generic expression for people who live on the streets (rough sleepers), people without permanent living arrangements, or those with inadequate habitations. In Hungary, according to the law, people experiencing homelessness are persons without any registered place of residence or whose registered place of residence is the accommodation for homeless individuals [1].

Although previous research has acknowledged the difficulty in the assessment of the scale of homelessness across Europe [2], it has been noted that the number of people experiencing homelessness is increasing in the European Union [3]; approximately 700,000 people are homeless on any given day, and this number has increased by 70% in the last 10 years [1]. In Hungary, systematic resources on homeless populations are scarce, meaning that there is a lack of basic demographic studies, and no public databases are available on the estimates of the size of the group.

Homelessness, Inequalities, and Health

The state of homelessness can be described as both a cause and a consequence of poor health status, social exclusion, and marginalization [2]. According to research, the health effects produced by homelessness include significantly higher rates of bacterial and viral infections, diabetes, hypertension, and cardiovascular disease compared with populations with adequate housing options [4]. Similar results emerged when looking at the life expectancy of homeless and general populations; on average, a decrease by 11 years for homeless men and 15 years for homeless women was measured [4].

Furthermore, earlier research suggests that despite the poor health status of homeless populations, health services designed for their treatment are often described as insufficient and limited in their accessibility, availability, and appropriateness [5]. An earlier study conducted in the United States also noted a medicalization process among homeless services and the practice of providing services for homeless individuals to conform them to specific behaviors [6]. As a result, underdiagnoses and undertreatment of health conditions are strongly prevalent [7,8], significantly underpinning the necessity to develop novel approaches and interventions to address health inequalities that have existed for decades, as such disparities lower life expectancy and strengthen social exclusion.

Digital Tools and Digital Inclusion as Potential New Approaches

The COVID-19 pandemic has accelerated the adoption of digital technologies in health care systems in many countries that experienced various types of lockdowns between 2020 and 2022. The World Health Organization's assessment of the European digital health landscape describes that during the COVID-19 pandemic, many digital health tools moved from being viewed as a potential opportunity to becoming an immediate necessity, and their use increased substantially [9]. The pandemic is also believed to have demonstrated that the lack of broadband access to the internet has an influence on the social determinants of health [10].

Although the expansion of the digital component of health care systems is considered a forward-looking development, it has raised accessibility issues for vulnerable strata, such as homeless populations. Physical barriers in the form of lack of access to technological equipment, as well as educational barriers in being unable to use the technology, may contribute to the inaccessibility of services and resources, further depriving a segment of the population that is already marginalized. This very possibility would negatively impact behaviors and stressors and might further contribute to poorer health outcomes for those who are digitally excluded, widening the already existent digital inequality landscape [11,12].

A systematic review analyzing studies from 2015 to 2021 with the research questions (1) "What mobile health–related technology is used by homeless populations?" and (2) "What is the health impact of mobile technology for homeless populations?" found that most homeless participants across the 17 studies included in the review owned a mobile phone or smartphone and 80% (1205/1507) owned a mobile phone. Age appeared to be a significant factor regarding ownership and use, and confirmatory responses to questions on access to mobile internet services, smartphone functions, and apps dropped significantly [11]. Heaslip et al [11] mentioned the lack of charging points, limited or no access to data traffic, and anxiety over potential theft and harassment as barriers to mobile phone use. Other barriers presented were privacy concerns and distrust in the management of data, tracking of information, the government, and the "system" [11]. Beyond physical barriers and trust issues, access to digital health might be hindered by the lack of skills required for their use. Populations at risk for limited health literacy, such as homeless populations [13], are similarly vulnerable to having challenges with digital tools [14]. Poor IT skills among homeless populations have been implicated in poor mental health outcomes [14].

However, despite existing barriers, several studies have reported the interest of the homeless population in digital health tools [11]. Atkins et al [15] noted that their study participants were positive about using a mobile phone to obtain advice and help address issues such as depression, anxiety, self-harm, abuse, substance use, emotional problems, insomnia, and stress. In all, 3 studies showed that interest in appointment and prescription reminders among homeless populations is prevalent [15-17].

Early Examples: Attitudes Toward Digital Health Among Homeless Individuals in Hungary

As the above literature review supports, physical barriers to accessing technologies and educational barriers in relation to digital technologies might strengthen the already existing digital inequalities to the detriment of homeless population, whereas the use of the internet was shown to be significantly associated with better self-rated health in older adults [18,19] and more favorable health behaviors concerning cancer prevention [20]. Studies conducted mainly in the United States, Canada, and the United Kingdom, focusing less on continental Europe or lower-income countries, suggest these findings [11].

The main aim of this study was to examine whether these assumptions are valid in the context of Hungarian homeless population and to suggest recommendations for public health policy makers. Thus, the main research questions were whether (1) homeless populations use digital tools for health-related reasons in Hungary and (2) clearly identifiable variables, such as the institutional and social services environment, age, education, or other demographic data can be associated with such use. In the case of social institutional characteristics, we assume that existing barriers and potentials of unique institutions to digital inclusion might be considered and offered as background information for potential interventions for digital inclusion, which we aim to examine as part of the second research question.

This study fits into a broader set of research undertaken by the joint action of the Digital Health Research Group at Semmelweis University and the Hungarian Charity Service of the Order of Malta (HCSOM), aiming to analyze the relationship between digital health and homeless populations in Hungary. Previous research has studied the attitudes of homeless individuals toward telecare services, with the main finding being that trust in the general health care system leads to trust in digital health solutions [12]. This study also served as an assessment tool for analyzing the viability of a telecare system planned to be launched by the HCSOM.

Methods

Participating Institutions

Homelessness can be categorized using different methods; Edgar et al [21] identified 6 different groups. As for the classification and definition of “homelessness” in this study, we decided to include all individuals who had engaged with institutions providing homeless services according to the categories of the European Typology of Homelessness and Housing Exclusion, the standard used by European Union member states for reporting on homelessness and precarious housing circumstances [22].

Altogether, 6 types of institutions providing social services for homeless populations participated in the study (Table 1). Although family shelters are not considered a part of the homeless social services according to the law in Hungary (these institutions are operated under the Child Protection Act), they were included in the study based on the housing instability of their clients and the temporary nature of the provided accommodation.

Table 1. List and characteristics of participating institutions and social services (N=662).

Type of service	ETHOS[a] classification	Client	Participating institutions (N=28), n (%)	Participants, n (%)
Street outreach service	1.1	Rough sleepers	4 (14)	106 (16)
Day shelter	N/A[b]	Homeless persons (no accommodation offered)	5 (17.9)	167 (25.2)
Night shelter	2.1	Homeless persons (accommodation offered only for short periods)	7 (25)	145 (21.9)
Temporary shelter	3.2-7.2	Homeless persons (accommodation offered for longer periods with a maximum of 1+1 years)	7 (25)	178 (26.8)
Temporary shelter with a focus on health improvement	3.2-7.2	Homeless persons with severe health status (accommodation offered for longer periods with a maximum of 1+1 years)	2 (7.1)	40 (6)
Family shelter	7.2	Homeless families (accommodation offered for longer periods with a maximum of 1+1 or 2 years)	3 (10.7)	48 (72.5)

[a]ETHOS: European Typology of Homelessness and Housing Exclusion.

[b]N/A: not applicable.

The Surveying Process

The research team formulated a questionnaire (Multimedia Appendix 1) based on the Digital Inclusion Survey used in a report by Pathway, the United Kingdom’s leading homeless health care charity [23]. The original questionnaire was translated to Hungarian by 2 independent medical translators, and their versions were merged by a consensus meeting. This Hungarian draft questionnaire was adapted to the local specialties during a workshop with social workers of the HCSOM. Before administering the questionnaire to a wider population, a test survey with 10 participants was completed to check its clarity and intelligibility. The selection of test group members was managed by one of the participating social establishments. To maximize the impact of the test survey, it was requested to use a diverse group of homeless clients with respect to gender, age, health status, and type of accommodation.

Subtle changes in wording were applied during the finalization of the survey material based on this feedback.

Between April 19, 2021, and August 11, 2021, the research group surveyed 662 people in Budapest, Hungary, with the cooperation of 28 institutions that provide various social services for homeless individuals. The respondents participated in the study on a voluntary basis. Our research team contacted the institutions, and their social workers asked homeless clients to fill out the questionnaires in a paper and pencil form. Social workers were allowed to help in the interpretation of questions but were not allowed to influence the answers. When a respondent was using multiple social services (eg, day and night shelter), we asked individuals to complete the questionnaire at the institution that provided the most relevant service for them to reduce duplicate responses.

The questionnaire enquired about sociodemographic data (age, gender, level of education, self-defined homelessness, and length of being homeless) and health status (frequency of medical visits, existing medical diagnoses, and self-assessment of health status). Questions 6-10 were used to gather information about health knowledge and general literacy skills, whereas questions 11-13 and 14-17 asked about access to mobile phones and the internet. Next, questions 18-21 inquired about internet use habits and questions 22 and 23 about potential barriers and enablers of internet access. Question 24 presented a set of statements about digital health literacy, and question 25 asked about mobile apps.

Reference Group

For the questions "How frequently do you visit a medical doctor/do you use medical services?" "Do you have any chronic disease or a long-term health problem?" "Have you ever used the Internet for any purpose? If yes, have you used it in the last six months?" and "Have you ever used any health-related mobile applications?" the responses of a representative sample of the Hungarian population were used as a reference group to provide more context. This representative survey was conducted by the Digital Health Working Group of Behavioral Institute of Semmelweis University between October 5, 2021, and October 13, 2021, and consisted of responses from 1500 Hungarian people in the framework of the "E-Patients in Hungary" study [24].

Statistical Analysis

As part of the quantitative analysis, we descriptively examined frequencies, averages, and percentage distributions. Use of technology and its various correlates (demographic variables and variables related to access to health services) were compared with a single variable analysis using Pearson chi-square test, with a significance level of $P<.05$.

In the multivariate analysis, a binary logistic regression model was used. The method was used to examine the background factors for the question "Have you ever used the internet for health reasons?" which is the dependent variable. The control variables were gender, type of institution and social service, level of education, age, frequency of medical visits, and prevalence of chronic illness. Independent variables affecting the dependent variables were selected using enter regression. The significance of the regression coefficients of the given variables was described using P value of the Wald. Variables with $P<.05$ were retained in the final model.

Data were analyzed using SPSS (version 26; IBM Corp) statistics software [25].

Ethics Approval

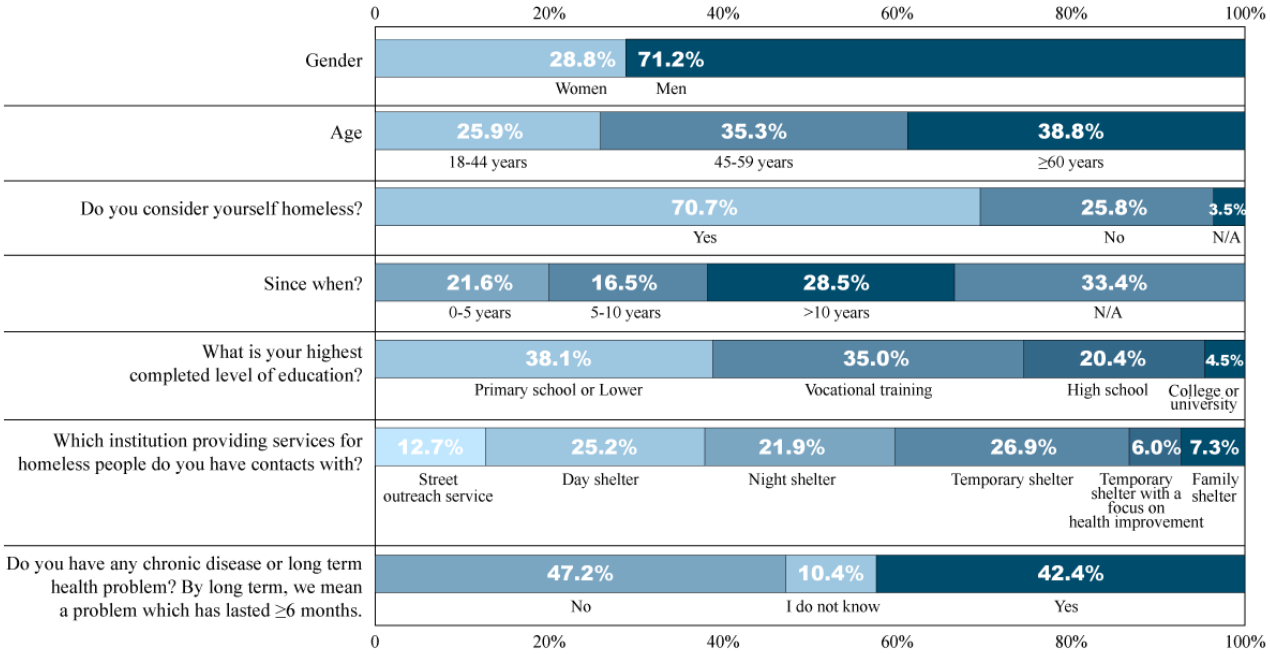
The data collection was anonymized. Written informed consent statements were obtained in all cases, and ethics approval for the study was issued under TUKEB:133/2020 and IV/10927/2020/EKU by the Scientific Research Ethics Committee of the Medical Research Council of Hungary.

Results

Demographics

The research group surveyed 662 adults in Budapest, Hungary, recruited from 28 social institutions providing services for people experiencing homelessness. Of the respondents, 71.2% (459/662) were men. Of the recruited participants, 38.8% (247/662) represented the age group of >60 years, whereas participants aged 18 to 44 years accounted for only 25.9% (165/662). The mean age was 53.9 years with an SD of 13.08 years. The majority, 70.7% (468/662), considered themselves homeless, whereas 25.8% (171/662) of the respondents did not consider themselves homeless. A total of 66.6% (441/662) of respondents also indicated how long they were experiencing homelessness: 21.6% (143/662) had been homeless for 1 to 5 years, 16.5% (109/662) for 5 to 10 years, and 28.5% (189/662) for >10 years, with a mean of 11.35 years and an SD of 9.27 years. Most of the respondents had only primary education (252/662, 38.1%) or vocational training (232/662, 35%), whereas 20.4% (135/662) of the respondents had graduated high school, and 4.5% (30/662) of the respondents said they had completed their college or university education. The key demographic parameters are shown in Figure 1.

Figure 1. Key demographics of the homeless group. N/A: not applicable.



Health Status

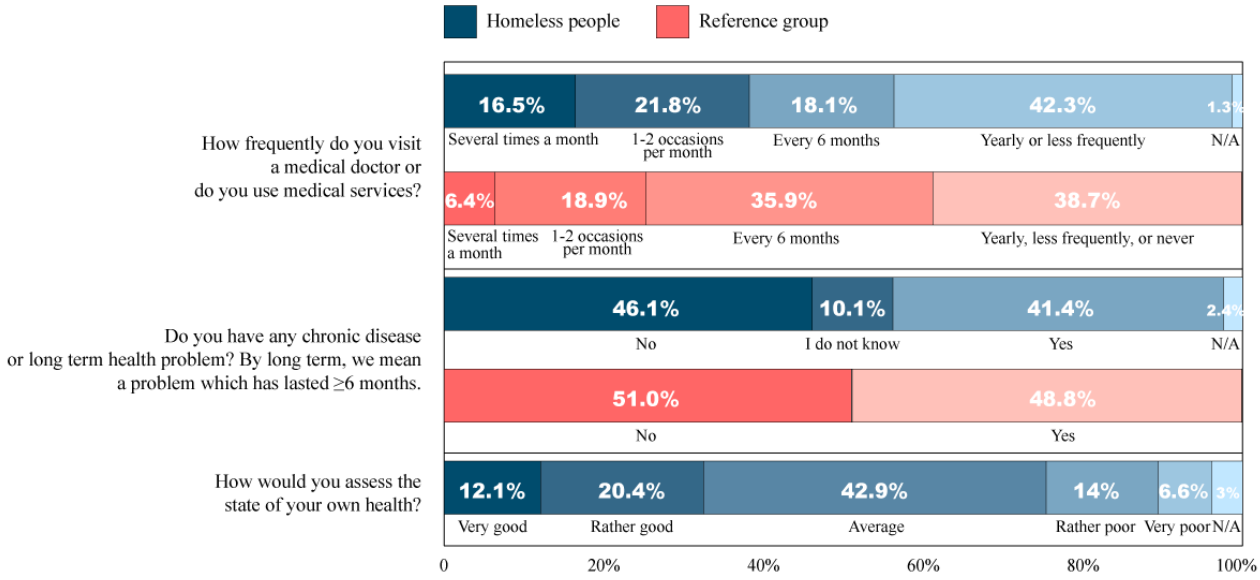
As key independent variables, we surveyed the health status of the respondents and compared them with the data of the reference group. A total of 16.5% (109/662) of the respondents said that they visited their physician or used health care services more than once a month, which was relatively frequent compared with the reference group, wherein 6.4% (96/1500) respondents said they visited their physician weekly, more than once a week, or more than once a month. Within the homeless group, 21.8% (144/662) of the respondents said they visited their physician every 1 or 2 months, which is almost the same as the result for the reference group (284/1500, 18.9%). The main difference was that most of the homeless group, 42.3% (280/662), visited their physician only yearly or less frequently, whereas 35.9% (539/1500) of the reference group said they used health care services 1 to 2 occasions per year, and only 13% (195/1500) of the respondents reported going to the physician's office yearly.

Of the homeless participants, 46.1% (305/662) reported no chronic diseases or long-term illnesses requiring treatment lasting for ≥6 months, but there was only a slight difference in the distribution of those who did (274/662, 41.4%). Those who had a chronic disease listed chronic obstructive pulmonary disease, asthma, diabetes, hypertension, mental illnesses, and chronic heart conditions among others. For the reference group, 48.8% (732/1500) of the respondents responded that they had a long-term illness, whereas 51% (765/1500) said that they did not have any.

Regarding the homeless group evaluating their own health, 12.1% (80/662) and 20.4% (135/662) of the respondents said "very good" or "rather good," respectively, whereas most people (284/662, 42.9%) considered it "average." In addition, 14% (93/662) and 6.6% (44/662) of the respondents said they considered their health "rather poor" and "very poor," respectively (Figure 2).

When asked about what channels they were using when informing about medical issues, 20.5% (136/662) of the respondents said they were searching for it on the web. This came in third after asking the primary care physician for information (352/662, 53.1%) and the social worker in the social institution (260/662, 39.2%), which meant they might have been consulting the internet for medical purposes more often than they asked their family members or friends (108/662, 16.3%).

Figure 2. Key demographics concerning health status of the homeless group. N/A: not applicable.



Access to Technology and Web-Based Services

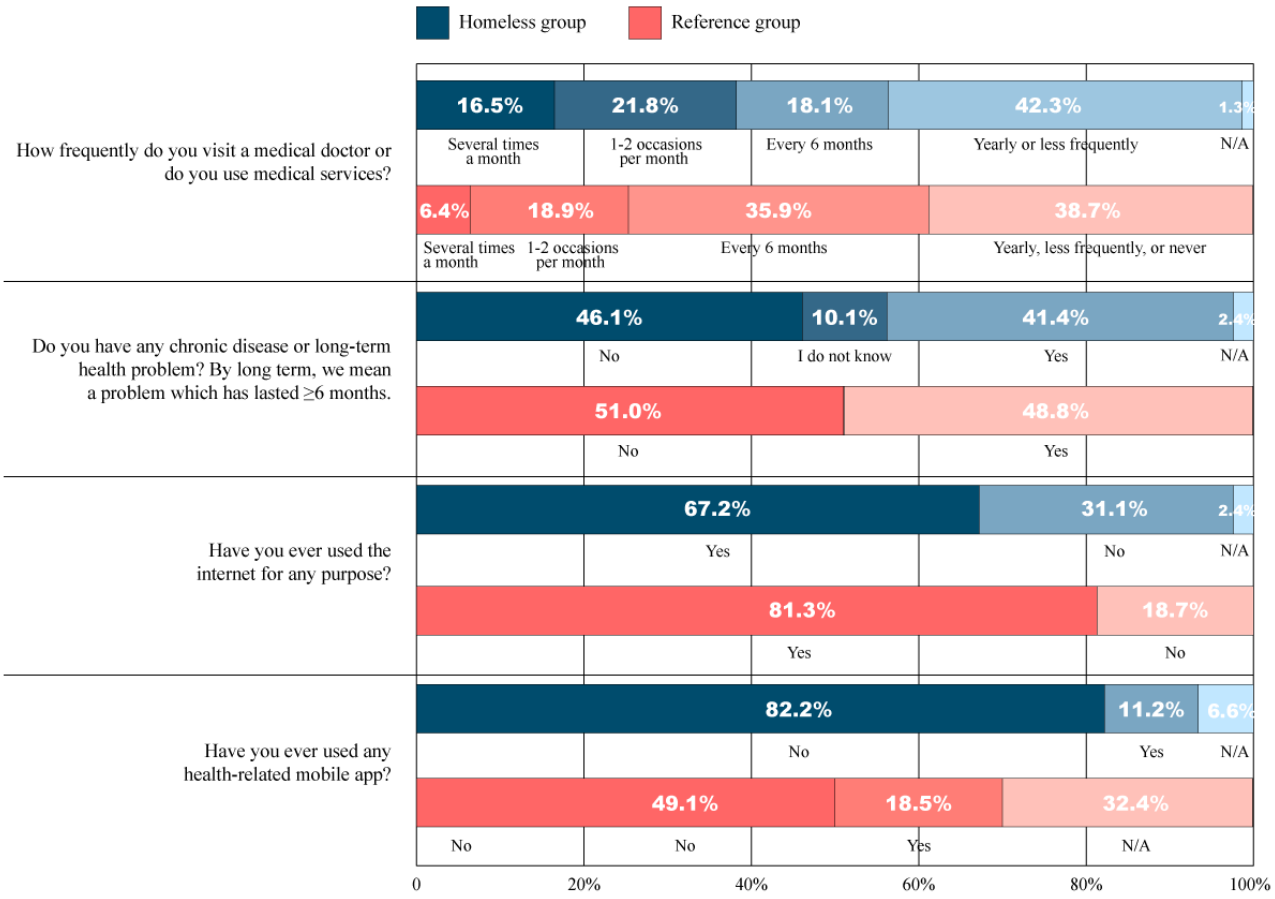
For the multiple-choice question, "How do you access the internet at the moment?" 98 people (98/551, 17.8%) said that they had their own smartphone with a data contract, 100 people (100/551, 18.1%) said that they had their own smartphone using a pay-as-you-go facility, 118 people (118/551, 21.4%) said that they had their own smartphone and accessed the internet via free Wi-Fi hotspots, 136 people (136/551, 24.7%) said that they accessed the internet through a publicly available PC in social institutions or shelters, only 15 people (15/551, 2.7%) said that they had their own PC, and 84 people (84/551, 15.2%) responded with "Other." In the latter category, answers included the use of other people's phones, "internet cafés," or ownership of a tablet, but a frequent response was that they had no means to access it, they did not care, or they did not use it. Only a few people access the internet in multiple ways (70/662, 10.6% in 2 ways, 12/662, 1.8% in 3 ways, and 4/662, 0.6% in 4 ways), while more than half of the respondents have access to it in only one way (359/662, 54.2%) or in no way (217/662, 32.8%).

In the reference group, 81.3% (1220/1500) of the respondents said that they used the internet frequently, whereas in the homeless group, 67.2% (445/662) of the responses were affirmative when asked if they ever used it for any purpose (Figure 3). Of those who used it, 52.9% (350/662) said they had used it in the past 6 months. However, daily use was significantly less, 34.6% (229/662), and an additional 10.6% (70/662) of the respondents said that they were using it more times a week. No correlation with age, type of institution and social service, gender, education, length of homelessness, or frequency of medical visits was found after cross-tabulation.

Most respondents of the homeless population (461/662, 69.6%) said that they owned a mobile phone. In addition, 39.9% (264/662) of the respondents also said that their mobile phone had a smartphone function, and 11.2% (74/662) of the respondents of the homeless group said that they had used at least one mobile health (mHealth) app, whereas this ratio was 18.5% (277/1500) in the reference group. In the homeless group, those who responded positively to the questions mentioned using apps for step counting, accessing emergency help, obtaining relevant medical information, and providing health data. mHealth apps were associated with 2 variables. Chi-square test results were significant for the type of institution and social service (P=.02) and frequency of medical visits (P=.03), meaning that mHealth apps were more frequently used in temporary shelters than in any other type of institution and social service, and with an increasing frequency of medical visits, the frequency of mHealth app use also increased.

For the question of how experienced they considered themselves when it came to internet use, 10% (66/662) of the respondents said "very much so," 14.5% (96/662) of the respondents said "rather experienced," and 21.5% (142/662) of the respondents said "mediocre," whereas 10.3% (68/662) of the respondents considered themselves "rather not experienced," and the most prevalent response, 35.3% (234/662), was "not at all" experienced. A total of 8.5% (56/662) did not respond to the question. When cross-tabulating self-reported technology literacy with age, education, gender, homelessness, type of institution and social service, and frequency of medical visits, chi-square tests were significant for age (P<.001), type of institution and social service (P=.01), and education (P=.01), meaning that with age, the level of self-reported technological literacy decreased, whereas with higher levels of education, self-identified technology literacy increased. Most of the respondents did not consider themselves as experienced technology users; this most significantly characterized the clients of temporary shelters with a focus on health improvement, whereas most experienced technology users made use of the social services of daily and family shelters.

Figure 3. Health and internet use characteristics of the homeless and reference groups. N/A: not applicable.



Barriers and Enablers of Internet Use

For the multiple-choice question, "What barriers, if any, restrict your internet use?" of the 682 responses, 210 (30.8%) said that nothing hindered it; 104 (15.2%) said there were not enough free Wi-Fi hotspots; only 46 (6.7%) said they had a smartphone, but they did not have a data contract or pay-as-you-go facility; and 52 (7.6%) said that they had internet access, but they did not know how to use the internet. Of the 682 responses, 146 (21.4%) said that they did not have a smartphone and 60 (8.8%) said that there were not enough publicly accessible PCs (eg, in institutions providing social services). In addition, of the 682 responses, 64 (9.4%) said that they could not access the internet anywhere.

For the question, "What would help you use the internet more?" of the 598 responses, 145 (24.2%) wished to have a smartphone, 110 (18.4%) responded better access (they had a smartphone but did not have an available internet connection option), another 56 (9.4%) also responded better access (they used PCs in institutions providing social services, but only a limited number of devices were available), 135 (22.6%) responded more knowledge (they did not know how to use the internet, and it would have helped if they could get assistance); however, for most people, 152 respondents (25.4%), the question was not relevant as they already used the internet as much as they wanted.

Health-Related Internet Use

For the question, "Have you ever used the internet for health reasons?" 34.6% (229/662) of the homeless population said that they did. In the reference group, 10.7% (160/1500) used it every day, 18.4% (276/1500) weekly, 18.2% (273/1500) monthly, and 24% (360/1500) less, encompassing 71.3% (1069/1500) of the representative sample. This means that the general population used the internet for medical purposes more than twice as frequently as the homeless population.

When cross-tabulating with gender, age, type of institution and social service, education, frequency of medical visits, and self-evaluation of health status, chi-square tests were significant for gender (P=.007), age (P<.001), and frequency of medical visits (P=.01), meaning that younger women respondents and those who went to the physician's office more frequently tended to use the internet more frequently for health-related issues.

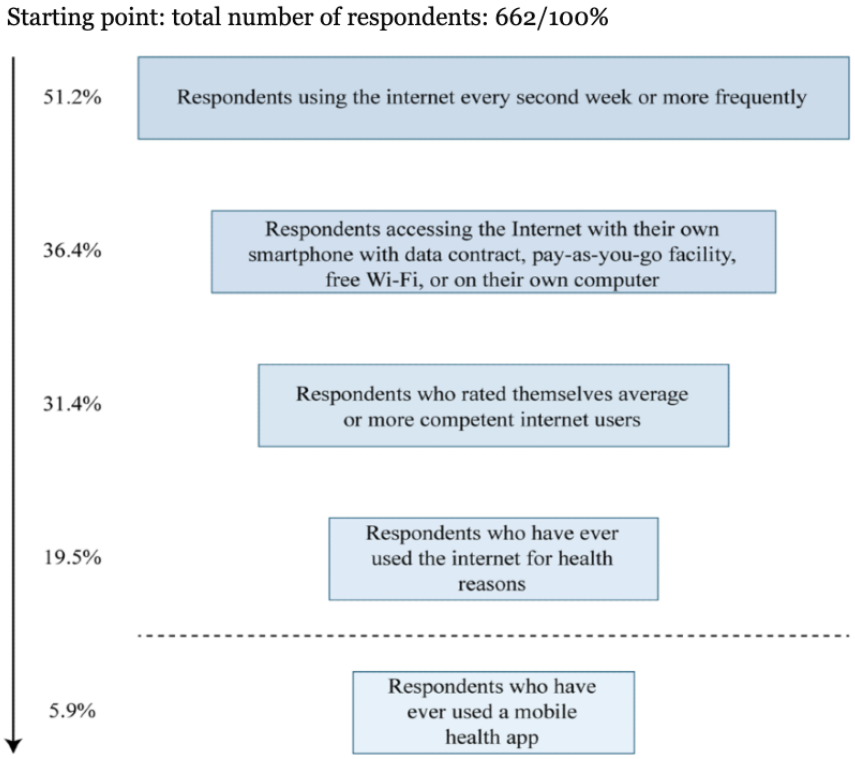
A Digitally Engaged Group of People Experiencing Homelessness

In the course of our analysis, we found a specific subpopulation in the sample identified as a "digitally engaged group of people experiencing homelessness." The members of this group were specific in the sense that they did not need further digital inclusion. This group was selected for further analysis based on the following inclusion criteria.

First, we selected respondents who said that they were using the internet at least every second week (339/662, 51.2%). In the

next step, we asked the respondents who reported smartphone ownership with data contract, pay-as-you-go facility, or free Wi-Fi or computer or tablet ownership to the question "How do you currently access the internet?" (241/662, 36.4%). We then filtered out the respondents who did not have a sense of being an average or more competent internet user (208/662, 31.4%). Furthermore, we selected those who responded "yes" to the question whether they had ever used the internet for health-related reasons (129/662, 19.5%). We also considered filtering the subpopulation based on the question "Have you ever used any health-related mobile application?" but as only 18.5% (277/1500) in the reference group responded positively to the question, we expected a significantly lower number in the homeless population, bordering analyzability. In contrast, the low number in the reference population indicates that mHealth app use is not necessarily meaningfully associated with overall health-related digital engagement. Thus, we created 2 subpopulations, a more broadly defined and a more strictly defined group, and analyzed their characteristics separately (Figure 4).

Figure 4. Flowchart for selecting the digitally engaged group of people experiencing homelessness.



When the selected subgroup included 19.5% (129/662) of the total homeless population, significantly more women were included in the subpopulation (47/129, 36.4%) than the original population (186/662, 28.8%). When cross-tabulating with gender, age, education, frequency of medical visits, prevalence of chronic illnesses, and type of institution and social service, chi-square test results were significant for the prevalence of chronic illness (P=.047); therefore, respondents with chronic illnesses were more likely to use the internet frequently for health-related reasons. Although the institutional setting was not an associative variable, temporary shelters (40/129, 31%) and day and night shelters (28/129, 21.7% and 22/129, 17%, respectively) housed most respondents in the subpopulation (90/129, 69.7%).

Of the 662 participants, we filtered out those who had never used a health-related mobile app (Figure 4). The selected subgroup included 5.9% (39/662) of the respondents of the total studied population. The gender ratio became balanced, which means that more women (14/39, 36%) were included in the subgroup than in the original population (186/662, 28.8%). When cross-tabulating with gender, age, education, frequency of medical visits, prevalence of chronic illnesses, and type of institution and social service, the chi-square test results were significant for the institutional setting (P=.03) and education (P=.04), which means that digital engagement of a homeless person tended to depend on the type of homeless shelter the respondent frequented, and respondents with higher levels of completed education tended to be more digitally engaged.

Multivariate Analysis

Chi-square test results showed that gender, age, and frequency of medical visits were associated with health-related internet use; however, to analyze which demographic or health status variables influenced health-related internet use, a binary logistic regression model was necessary.

The dependent variable was health-related internet use, and we entered gender (1=woman and 2=men), age (as a continuous variable), type of institution and social service (6 categories), education (4 categories), frequency of medical visits, and the prevalence of chronic conditions in the model.

The logistic regression model was found to be significant (Nagelkerke R^2=0.154). After controlling for all the abovementioned variables, we found that health-related internet use showed a strong dependency on age and a statistically

significant association with gender, level of education, and the prevalence of chronic conditions (P<.05). This means that younger homeless women who did not have any chronic conditions tended to use the internet more for health-related issues (Table 2).

Table 2. Results of the logistic regression model (Nagelkerke R^2=0.154)[a].

	B (SE)	Wald test (*df*)	*P* value	Exp (B)
Gender (1=female; 2=male)	−0.480 (0.222)	4.660 (1)	.03	0.619
What is your highest completed level of education?	__[b]	9.186 (3)	.03	—
What is your highest completed level of education? (1=primary school)	0.458 (0.483)	0.899 (1)	.34	1.581
What is your highest completed level of education? (2=vocational training)	−0.191 (0.480)	0.158 (1)	.69	0.826
What is your highest completed level of education? (3=high school)	−0.141 (0.495)	0.081 (1)	.78	0.869
How frequently do you visit a medical doctor or do you use medical services?	0.155 (0.099)	2.453 (1)	.12	1.168
Do you have any chronic disease or a long-term health problem? By long-term, we mean a problem which has lasted six months or longer.	−0.481 (0.238)	4.077 (1)	.04	0.618
Age	0.049 (0.009)	30.033 (1)	<.001	1.050
Which institution providing services for homeless people do you have contacts with?	—	3.607 (5)	—	—
Which institution providing services for homeless people do you have contacts with? (1=outreach service)	0.606 (0.458)	1.752 (1)	.19	1.833
Which institution providing services for homeless people do you have contacts with? (2=day shelter)	0.356 (0.397)	0.804 (1)	.37	1.428
Which institution providing services for homeless people do you have contacts with? (3=night shelter)	0.058 (0.431)	0.018 (1)	.89	1.059
Which institution providing services for homeless people do you have contacts with? (4=temporary shelter)	0.109 (0.434)	0.063 (1)	.80	1.115
Which institution providing services for homeless people do you have contacts with? (6=family shelter)	0.223 (0.585)	0.145 (1)	.70	1.249
Constant	−2.052 (0.838)	6.002 (1)	.01	0.128

[a]Dependent variable: Do you ever use the Internet for health reasons? (0=no; 1=yes).

[b]Not available.

Discussion

Principal Findings

Homeless adults experience an early onset of geriatric conditions, a complex set of chronic diseases, and premature mortality [26,27], as their access to adequate health care services is generally poor. Such disparities lower life expectancy and strengthen social exclusion. To mitigate health inequalities among homeless populations, digital technology [12], a new health determinant, can be considered on a broader scale. In a previous study by the Digital Health Research Group [12] at Semmelweis University that examined the attitudes and openness of homeless individuals regarding telecare in a Hungarian sample, a significant fraction of people experiencing homelessness with mid- or long-term residency in homeless shelters was open to the use of telecare via live web-based video consultation. As a step forward in assessing the feasibility of launching a comprehensive telehealth project and disseminating other well-being programs, the research team conducted this survey assessing existing access to digital platforms (smartphones and internet) and barriers in both physical and educational spaces among homeless populations.

On the basis of our findings, the surveyed homeless population showed an aptitude toward health-related technology use and had partial access to digital tools. Overall, the results respond to our first research question positively, that is, homeless populations use digital tools for health-related reasons.

A significant proportion of respondents had a mobile phone (461/662, 69.6%), and a lower but still significant number of respondents possessed a smartphone (264/662, 39.9%). These findings are congruent with the results presented in the literature, although according to our findings, the ownership of devices and access to the internet lag behind that of Western countries. In 2013, McInnes et al [28], in a systematic review, found that mobile phone ownership ranged from 44% to 62%, computer ownership from 24% to 40%, computer access and use from 47% to 55%, and internet use from 19% to 84% in this population. In 2017, Rhoades et al [29] found that the vast majority of homeless individuals (94%) owned a cell phone, more than half owned a smartphone, and 51% accessed the internet on their cell phones. One-third of the participants reported no internet use in the past 3 months [29]. In 2021, Thurman et al [30] analyzed feasibility studies related to mHealth interventions among people experiencing homelessness and found that 52% of the participants (n=31) reported having a personal cell phone, and of those with phones at baseline, the

majority (87%) reported that their phones were capable of SMS text messaging, picture messaging, and mobile app use.

Our results showed that people experiencing homelessness turn to their family physician and social workers the most frequently for help with medical issues, but their third most frequent choice is the internet (20.5%), even before asking family members or friends. In total, 34.6% (229/662) of the respondents said they had used the internet for medical purposes, and 11.2% (74/662) of the respondents had already used a medical mobile app.

In addition, we have to consider technological limitations. The first iPhone was launched in 2007, which introduced the concept of smartphones, the spread of smartphone-based internet use, and personalized web-based searches. Technological adoption is slower in lower socioeconomic groups, and previous studies found that rates of smartphone and internet use among homeless populations were lower than those among housed, low-income adults of any age [31], which might explain the generally lower internet use statistics for this specific group. This is in line with the findings of Von Holtz et al [32] showing that, while experiencing homelessness, participants experienced a 68% less likelihood to access the internet than when they were housed; however, our main results show that the idea of involving homeless populations in the digital health ecosystem can already be based on solid use patterns, which can be further extended.

Age as a Key Predictor of Health-Related Internet Use

On the basis of our findings, the response to our second research question, that is, clearly identifiable variables, above all institutions and social services, and beyond that, age, education, or other demographic data can be associated with health-related internet use, had to be partially rejected. Neither chi-square tests nor the binary regression model showed statistically significant results. The type of institutional access and social services provided did not relate to access and use of digital tools and the internet, except for the digitally engaged subgroup. In contrast, our logistic regression model showed that age, gender, level of education, and prevalence of chronic conditions are variables that statistically significantly influence health-related internet use.

In line with our results, Harris et al [33] found age to be a key sociodemographic variable affecting the use of technology by homeless individuals. The participants of that study felt that the shift in the United Kingdom to more digital social services had assumed that users were well versed with IT, although this may not be the case.

Although age seemed not to play a key factor in homeless individuals accessing technology, as most of the respondents had a mobile phone (461/662, 69.6%), mostly representing the age group of >60 years, it might be a crucial factor when it comes to their own perception of competence in using web-based services and health-related internet use. Younger respondents (age group 18-44 years) considered themselves rather competent, whereas older respondents (age groups 45-59 years and >60 years) did rather not or did not at all consider themselves competent when it came to using the internet. Moreover, the regression model showed that the younger a homeless respondent was, the more likely they were to use the internet for health-related reasons.

Gender, Level of Education, and Prevalence of Chronic Conditions

The regression model showed that gender was an explanatory factor when it came to health-related internet use, which means that women in the homeless group tended to use digital tools mainly for health-related purposes. This is congruent with the trends in the general population, as Resch et al [34] found that women were more engaged in using the internet to search health-related information in Germany (n=1006), and Rising et al [35] through the 2017 and 2018 National Cancer Institute Health Information National Trends Survey (n=6789) found that in the United States, women were more likely than men to use digital health tools. As a noteworthy limitation, it has to be mentioned that women were almost 2.5 times more underrepresented in the sample (186/662, 28.8%), which might have influenced mHealth use patterns along gender lines.

Regarding the level of education, those who had completed higher levels of education were more inclined to use digital health tools, although only 4.5% (30/662) of the sample said they had completed college or university education, which, similar to the gender composition of the sample, might influence use patterns. In contrast, this finding is congruent with the self-assessment of technological literacy. Chi-square test results were significant for education (P=.01) when cross-tabulating with self-assessment of digital competencies, meaning that with higher levels of education, the sense of technology literacy increases, which might result in more frequent use.

Concerning the prevalence of chronic conditions, the results showed that homeless individuals without chronic diseases or any long-term illnesses tended to use the internet more for health-related purposes, which might originate from the pattern that those who were more concerned about their own health tended to use a diverse tool kit for health care and well-being, including digital tools, whereas those with serious chronic illnesses might tend to neglect their state because of their struggle to accommodate basic human needs or lack of resources for accessing care [36].

Overall, the results of the regression model were in line with trends in the general population: younger and more educated people tend to use digital health tools [37,38], and this finding means that in the course of planning health care interventions for homeless populations, patterns observed in the general population might be taken as a base for further action.

Digitally Engaged Homeless Subpopulation

The homeless population was a diverse group in terms of health-related internet use and access to digital tools, with a significant number of digitally engaged participants. When analyzing the data, the research team found 2 broadly interpretable digitally engaged homeless subpopulations: a subpopulation without health-related mobile app use (129/662, 19.5%) and another with such use (39/662, 5.9%). Generally speaking, both digitally engaged groups included more women and younger respondents than the homeless population, which was in line with the findings of the regression model. The overall

results were also congruent with previous literature stating that low-income populations rely on smartphones rather than computers for internet access; the latter was less frequent than owning a smartphone in our sample as well [31].

A chi-square test on the association between demographic factors and the more broadly defined subgroup showed that the type of institution and social service as well as the level of education—the higher the level of completed education, the more substantial digital engagement—mattered as factors for becoming digitally engaged. Temporary shelters (40/129, 31%) and day and night shelters (28/129, 21.7% and 22/129, 17%, respectively) housed most respondents in the subpopulation (90/129, 69.7%), which means that long-term living conditions seem to be associated with digital inclusion. The same pattern emerged in the more strictly defined subgroup; a chi-square test on the association between demographic factors showed that only the type of institution providing social services mattered as a factor for becoming digitally engaged. Almost half of the selected subgroup used temporary shelters, whereas very few digitally engaged users were found among rough sleepers and those who used emergency accommodations.

Barriers and Enablers of Internet Use

Rice et al [39] reported that mobile phones can facilitate communication with family or friends and provide social support, which in turn has been shown to be associated with more favorable health outcomes [40]. In contrast, two-thirds of the participants of a cohort of 350 adults experiencing homelessness aged >50 years in Oakland, California, reported using their phones to communicate with their health care providers, suggesting both interest and feasibility [31].

However, several studies have shown homeless population's interactions with technology to be significantly affected by lack of resources and the structural constraints [33], which was also shown by our results. As the main barriers to accessing technology, respondents mentioned affordability of digital tools or data contracts, the low number of free Wi-Fi hotspots, and PCs available at social institutions. To foster internet use, a significant number of respondents suggested overcoming these barriers rather than urging the need for educational assistance.

In line with previous studies, in the context of homeless populations in Hungary, increasing public access to high-speed internet and providing discounted smartphones for high-need, low-income individuals may also increase access to the internet [41]. Moreover, Budapest lacks an adequate number of free Wi-Fi hotspots, and thus needs more of such hotspots installed [42]. As Raven et al [31] noted, private sector technology and telecommunication companies might also be incentivized to fund initiatives that increase the use of their services among underserved populations, thereby increasing access to reliable mobile technology.

Strengths

Studies examining health and technology-related behaviors in homeless populations tended to be conducted predominantly in the United States and Canada compared with little examination of the use of technology of homeless populations in other countries [11]. Thus, as Heaslip et al [11] also noted, further research is needed in the United Kingdom, Europe, and lower-income countries. This study aims to fill that gap by examining the accessibility and use of health-related technology in Central and Eastern Europe, more specifically in Hungary.

Compared with other studies that examine homeless populations in specific areas, the sample size of this study (N=662) is considered notable and large enough to draw statistically significant conclusions.

Limitations

The study sample represents urban homeless populations from Budapest, Hungary, where socioeconomic conditions might differ from those living in the countryside. Homeless population recruited in our study had a connection to the social infrastructure; therefore, rough sleepers and other people who were not connected to any social initiatives were not represented.

The research team relied exclusively on self-reporting of mobile phone ownership, internet access, and internet use and did not attempt in any way to verify these reports (eg, via phone bills, direct observation, or other methods).

Conclusions

Although health-related internet use statistics are lower than those in the general population, the results showed that the pattern of use is similar. The idea of involving homeless populations in Hungary in the digital health ecosystem is not far-fetched, but a rather viable concept, especially if barriers to access are systematically reduced and the enablers of use strengthened.

During the development of a digital ecosystem, several factors might be considered, such as the role of the institutions providing social and medical services. From an infrastructural point of view, the unavailability and poor affordability of devices and subscriptions and the lack of publicly available free Wi-Fi hotspots were mentioned as barriers to digital technological access. All these factors might be improved by making adequate changes, enabling more Wi-Fi hotspots and installing more publicly available computers in social institutions. In addition, an internet service scheme specifically designed for the homeless population (eg, prepaid services available for medical purposes) could facilitate a shift toward better digital health.

It is important to note that despite all the barriers to accessing digital technologies, our research identified a digitally engaged homeless subgroup, whose members are actively using digital tools for health purposes. With a deeper analysis of this group, characteristics, motivations, and potentials for widening access and use could be delineated, and this group could form a baseline for holistic and appropriate digital public health interventions.

Our preliminary analysis in this group already showed that the characteristics of accommodation also play a role in assessing the accessibility of homeless populations to digital health services. People experiencing homelessness with a more stable housing solution tend to be more open to digital technology and have more access to their own digital resources than others with less stable conditions. This information might be fruitfully used

when planning further complex and holistic digital health programs for homeless populations centered on institutions as already available resources for further development.

Acknowledgments

The research project was supported by the National Research, Development and Innovation Office (OTKA-FK 134372). The authors would like to thank the following colleagues for their support in the research: Róbert Jónás Tóth, Emília Morva, Tamás Szentkereszty, and all colleagues from the participating social institutions involved in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital health access and literacy survey for people experiencing homelessness.
[DOCX File , 12 KB - jmir_v24i10e38729_app1.docx]

References

1. Nagy-Borsy E, VágiZ Z, Skerlecz P, Szeitl B, Kiss I, Rákosy Z. Health status and health behaviour of the Hungarian homeless people. Arch Public Health 2021 Feb 02;79(1):15 [FREE Full text] [doi: 10.1186/s13690-021-00534-2] [Medline: 33531045]
2. Lech S, Schenk L, De la Torre Castro J, Schindel D. A retrospective analysis of the health and social situation of homeless people in Berlin: study protocol. Arch Public Health 2021 Mar 06;79(1):28 [FREE Full text] [doi: 10.1186/s13690-021-00546-y] [Medline: 33676557]
3. Fifth overview of Housing Exclusion in Europe. FEANTSA. 2020 Jul 23. URL: https://tinyurl.com/55uu8tne [accessed 2021-03-30]
4. Graffy P, McKinnon S, Lee G, Remington P. Life outside: a narrative ethnographic inquiry into the determinants of homelessness. J Poverty 2018 Dec 13;23(3):202-228. [doi: 10.1080/10875549.2018.1550133]
5. Third overview of housing exclusion in Europe 2018. Feantsa. 2018. URL: https://www.feantsa.org/download/full-report-en1029873431323901915.pdf [accessed 2021-03-30]
6. Lyon-Callo V. Medicalizing homelessness: the production of self-blame and self-governing within homeless shelters. Med Anthropol Q 2000 Sep;14(3):328-345. [doi: 10.1525/maq.2000.14.3.328] [Medline: 11036582]
7. Gunner E, Chandan SK, Marwick S, Saunders K, Burwood S, Yahyouche A, et al. Provision and accessibility of primary healthcare services for people who are homeless: a qualitative study of patient perspectives in the UK. Br J Gen Pract 2019 Aug;69(685):e526-e536 [FREE Full text] [doi: 10.3399/bjgp19X704633] [Medline: 31307999]
8. Jagpal P, Saunders K, Plahe G, Russell S, Barnes N, Lowrie R, et al. Research priorities in healthcare of persons experiencing homelessness: outcomes of a national multi-disciplinary stakeholder discussion in the United Kingdom. Int J Equity Health 2020 Jun 05;19(1):86 [FREE Full text] [doi: 10.1186/s12939-020-01206-3] [Medline: 32503544]
9. Use of digital health tools in Europe: before, during and after COVID-19. World Health Organization. 2021. URL: https://apps.who.int/iris/handle/10665/345091 [accessed 2022-01-15]
10. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]
11. Heaslip V, Richer S, Simkhada B, Dogan H, Green S. Use of technology to promote health and wellbeing of people who are homeless: a systematic review. Int J Environ Res Public Health 2021 Jun 25;18(13):6845 [FREE Full text] [doi: 10.3390/ijerph18136845] [Medline: 34202242]
12. Győrffy Z, Békási S, Döbrössy B, Bognár VK, Radó N, Morva E, et al. Exploratory attitude survey of homeless persons regarding telecare services in shelters providing mid- and long-term accommodation: the importance of trust. PLoS One 2022 Jan 6;17(1):e0261145 [FREE Full text] [doi: 10.1371/journal.pone.0261145] [Medline: 34990458]
13. Odoh C, Vidrine JI, Businelle MS, Kendzor DE, Agrawal P, Reitzel LR. Health literacy and self-rated health among homeless adults. Health Behav Res 2019 Oct;2(4):13 [FREE Full text] [doi: 10.4148/2572-1836.1055] [Medline: 34164607]
14. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
15. Adkins EC, Zalta AK, Boley RA, Glover A, Karnik NS, Schueller SM. Exploring the potential of technology-based mental health services for homeless youth: a qualitative study. Psychol Serv 2017 May;14(2):238-245 [FREE Full text] [doi: 10.1037/ser0000120] [Medline: 28481610]
16. Buccieri K, Molleson G. Empowering homeless youth: building capacity through the development of mobile technology. J Community Pract 2015 May 18;23(2):238-254. [doi: 10.1080/10705422.2015.1027802]

17. Asgary R, Sckell B, Alcabes A, Naderi R, Adongo P, Ogedegbe G. Perceptions, attitudes, and experience regarding mHealth among homeless persons in New York city shelters. J Health Commun 2015;20(12):1473-1480 [FREE Full text] [doi: 10.1080/10810730.2015.1033117] [Medline: 26313765]
18. Duplaga M. The association between internet use and health-related outcomes in older adults and the elderly: a cross-sectional study. BMC Med Inform Decis Mak 2021 May 06;21(1):150 [FREE Full text] [doi: 10.1186/s12911-021-01500-2] [Medline: 33957909]
19. Falk Erhag H, Ahlner F, Rydberg Sterner T, Skoog I, Bergström A. Internet use and self-rated health among Swedish 70-year-olds: a cross-sectional study. BMC Geriatr 2019 Dec 23;19(1):365 [FREE Full text] [doi: 10.1186/s12877-019-1392-8] [Medline: 31870313]
20. Xavier AJ, d'Orsi E, Wardle J, Demakakos P, Smith SG, von Wagner C. Internet use and cancer-preventive behaviors in older adults: findings from a longitudinal cohort study. Cancer Epidemiol Biomarkers Prev 2013 Nov;22(11):2066-2074. [doi: 10.1158/1055-9965.EPI-13-0542] [Medline: 24148971]
21. Edgar B, Harrison M, Watson P, Busch-Geertsema V. Measurement of homelessness at European Union level. European Commission Employment, Social Affairs and Equal Opportunities DG. 2007 Jan. URL: https://ec.europa.eu/employment_social/social_inclusion/docs/2007/study_homelessness_en.pdf [accessed 2022-03-25]
22. Amore K, Baker M, Howden-Chapman P. The ETHOS definition and classification of homelessness: an analysis. Eur J Homelessness 2011 Dec;5(2):19 [FREE Full text]
23. Dorney-Smith S, Burridge S, Bell J, Ellis J, Snowball L. Digital health inclusion for people who have experienced homelessness – is this a realistic aspiration? Pathway. 2017. URL: https://www.pathway.org.uk/wp-content/uploads/2013/05/Digital-Inclusion-Report-Final-for-web.pdf [accessed 2021-03-29]
24. Zsuzsa GE. Bence Döbrössy. ePatients in Hungary, Digital Health Solutions in Medicine. Budapest, Hungary: Semmelweis University, Group DHR; 2022.
25. IBM Corp. IBM SPSS Statistics for Windows, Version 26.0. Armonk, New York: IBM Corp; 2019.
26. Sudore RL, Cuervo IA, Tieu L, Guzman D, Kaplan LM, Kushel M. Advance care planning for older homeless-experienced adults: results from the health outcomes of people experiencing homelessness in older middle age study. J Am Geriatr Soc 2018 Jul;66(6):1068-1074 [FREE Full text] [doi: 10.1111/jgs.15417] [Medline: 29741765]
27. Baggett TP, Hwang SW, O'Connell JJ, Porneala BC, Stringfellow EJ, Orav EJ, et al. Mortality among homeless adults in Boston: shifts in causes of death over a 15-year period. JAMA Intern Med 2013 Feb 11;173(3):189-195 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1604] [Medline: 23318302]
28. McInnes DK, Li AE, Hogan TP. Opportunities for engaging low-income, vulnerable populations in health care: a systematic review of homeless persons' access to and use of information technologies. Am J Public Health 2013 Dec;103 Suppl 2:e11-e24. [doi: 10.2105/AJPH.2013.301623] [Medline: 24148036]
29. Rhoades H, Wenzel SL, Rice E, Winetrobe H, Henwood B. No digital divide? Technology use among homeless adults. J Soc Distress Homeless 2017 Mar 22;26(1):73-77 [FREE Full text] [doi: 10.1080/10530789.2017.1305140] [Medline: 31097900]
30. Thurman W, Semwal M, Moczygemba LR, Hilbelink M. Smartphone technology to empower people experiencing homelessness: secondary analysis. J Med Internet Res 2021 Sep 29;23(9):e27787 [FREE Full text] [doi: 10.2196/27787] [Medline: 34586073]
31. Raven MC, Kaplan LM, Rosenberg M, Tieu L, Guzman D, Kushel M. Mobile phone, computer, and internet use among older homeless adults: results from the HOPE HOME cohort study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10049 [FREE Full text] [doi: 10.2196/10049] [Medline: 30530464]
32. VonHoltz LA, Frasso R, Golinkoff JM, Lozano AJ, Hanlon A, Dowshen N. Internet and social media access among youth experiencing homelessness: mixed-methods study. J Med Internet Res 2018 May 22;20(5):e184 [FREE Full text] [doi: 10.2196/jmir.9306] [Medline: 29789281]
33. Harris J. The digitization of advice and welfare benefits services: re-imagining the homeless user. Housing Stud 2019 May 09;35(1):143-162. [doi: 10.1080/02673037.2019.1594709]
34. 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]
35. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the us population by patterns of mobile health use for health and behavioral tracking: analysis of the national cancer institute's health information national trends survey data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]
36. Omerov P, Craftman AG, Mattsson E, Klarare A. Homeless persons' experiences of health- and social care: a systematic integrative review. Health Soc Care Community 2020 Jan;28(1):1-11. [doi: 10.1111/hsc.12857] [Medline: 31524327]
37. Bol N, Helberger N, Weert J. Differences in mobile health app use: A source of new digital inequalities? The Information Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
38. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]

39. Rice E, Kurzban S, Ray D. Homeless but connected: the role of heterogeneous social network ties and social networking technology in the mental health outcomes of street-living adolescents. Community Ment Health J 2012 Dec;48(6):692-698 [FREE Full text] [doi: 10.1007/s10597-011-9462-1] [Medline: 22075769]
40. Uchino BN, Cacioppo JT, Kiecolt-Glaser JK. The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health. Psychol Bull 1996 May;119(3):488-531. [doi: 10.1037/0033-2909.119.3.488] [Medline: 8668748]
41. Jaeger PT, Bertot JC, Thompson KM, Katz SM, DeCoster EJ. The intersection of public policy and public access: digital divides, digital literacy, digital inclusion, and public libraries. Public Library Quarterly 2012 Jan;31(1):1-20. [doi: 10.1080/01616846.2012.654728]
42. WiGLE homepage. WiGLE. URL: https://wigle.net/map?maplat=47.4999887892398&maplon=19.055079999999705&mapzoom=6&coloring=density [accessed 2022-10-13]

Abbreviations

HCSOM: Hungarian Charity Service of the Order of Malta
mHealth: mobile health

Edited by A Mavragani; submitted 13.04.22; peer-reviewed by L VonHoltz, N Karnik; comments to author 16.06.22; revised version received 16.08.22; accepted 22.08.22; published 19.10.22.

Please cite as:
Radó N, Girasek E, Békási S, Győrffy Z
Digital Technology Access and Health-Related Internet Use Among People Experiencing Homelessness in Hungary: Quantitative Survey
J Med Internet Res 2022;24(10):e38729
URL: https://www.jmir.org/2022/10/e38729
doi:10.2196/38729
PMID:36260379

Original Paper

General Health Statuses as Indicators of Digital Inequality and the Moderating Effects of Age and Education: Cross-sectional Study

Alexander J A M van Deursen[1], Prof Dr

Department of Communication Science, University of Twente, Enschede, Netherlands

Corresponding Author:
Alexander J A M van Deursen, Prof Dr
Department of Communication Science
University of Twente
Drienerlolaan 5
Enschede, 7500AE
Netherlands
Phone: 31 622942142
Email: a.j.a.m.vandeursen@utwente.nl

Abstract

Background: Considerable effort has been directed to offering online health information and services aimed at the general population. Such efforts potentially support people to obtain improved health outcomes. However, when health information and services are moved online, issues of equality need to be considered. In this study, we focus on the general population and take as a point of departure how health statuses (physical functioning, social functioning, mental health, perceived health, and physical pain) are linked to internet access (spanning internet attitude, material access, internet skills, and health-related internet use).

Objective: This study aims to reveal to what extent (1) internet access is important for online health outcomes, (2) different health statuses are important for obtaining internet access and outcomes, and (3) age and education moderate the contribution of health statuses to internet access.

Methods: A sequence of 2 online surveys drawing upon a sample collected in the Netherlands was used, and a data set with 1730 respondents over the age of 18 years was obtained.

Results: Internet attitude contributes positively to material access, internet skills, and health outcomes and negatively to health-related internet use. Material access contributes positively to internet skills and health-related internet use and outcomes. Internet skills contribute positively to health-related internet use and outcomes. Physical functioning contributes positively to internet attitude, material access, and internet skills but negatively to internet health use. Social functioning contributes negatively to internet attitude and positively to internet skills and internet health use. Mental health contributes positively to internet attitude and negatively to material access and internet health use. Perceived health positively contributes to material access, internet skills, and internet health use. Physical pain contributes positively to internet attitude and material access and indirectly to internet skills and internet health use. Finally, most contributions are moderated by age (<65 and ≥65 years) and education (low and high).

Conclusions: To make online health care attainable for the general population, interventions should focus simultaneously on internet attitude, material access, internet skills, and internet health use. However, issues of equality need to be considered. In this respect, digital inequality research benefits from considering health as a predictor of all 4 access stages. Furthermore, studies should go beyond single self-reported measures of health. Physical functioning, social functioning, mental health, perceived health, and physical pain all show unique contributions to the different internet access stages. Further complicating this issue, online health-related interventions for people with different health statuses should also consider age and the educational level of attainment.

(J Med Internet Res 2022;24(10):e37845) doi:10.2196/37845

KEYWORDS

digital inequality; health; MOS; eHealth; digital health; online health; age; education; survey; digital divide; attitude; health outcome; patient outcome; internet access; internet skill; technology skill

Introduction

Background

The World Health Organization (WHO) stresses that public health is an important topic on policy agendas in most Western countries. Considerable effort is directed to offering health information and services aimed at the general population online. Such efforts potentially support people in improved outcomes regarding their knowledge of health issues, health communication with professionals, decision-making about health issues, proper use of health services, and improved ways of taking care of themselves [1-3]. However, when health information and services are moved online, issues of equality need to be considered. Online information and services can also disempower marginalized people by violating their rights and autonomy [4], further entrenching their position. Digital inequality research typically considers how specific populations can benefit from access to online services and has shown that those most likely to experience health-related issues are also less likely to benefit from the internet in general [5]. In this respect, most attention has focused on, for example, age, racial and ethnic, and socioeconomic differences in access to online health. Actual well-being in terms of personal health is far less studied as a determinant of internet access in digital inequality research [5]. When considered, it is often simplified in binary terms or by a single self-rated health scale. In this study, we focus on the general population and take as a point of departure the way people with different health statuses—pertaining to general functioning and well-being—use the internet to obtain positive health outcomes, for example, in determining a medical condition from which one might suffer or making better health-related decisions. We attempt to provide an in-depth picture by focusing on different health statuses in relation to stages of internet access and online health outcomes. The paper is structured around 3 goals: to reveal to what extent (1) internet access (spanning internet attitude, material access, internet skills, and internet health use) is important for online health outcomes, (2) different health statuses (physical functioning, social functioning, mental health, perceived health, and physical pain) are important for obtaining internet access and outcomes, and (3) age and educational differences moderate the contribution of health statuses to internet access.

Internet Access and Outcomes

Resources and appropriation theory considers internet access as a process of appropriation following attitude, material access, skills, and use [6]. A positive attitude toward the internet is a first step toward using online health information and services [6]. Subsequently, material access involves having an internet connection and the required devices that provide internet access, such as desktops, laptops, tablets, and smartphones [6,7]. With the rapid increase in internet connections in Western countries, differences in materials (variety and quality of devices) are increasingly the topic of attention in this stage [7]. The required skills to use the internet range from operational skills (basic operations to use the internet) to information navigation (find, select, and evaluate sources of online information), communication (use online communication and interactions to understand and exchange meaning and acquire social capital), and content creation skills (create different types of quality content) [8]. The final access type in the current context involves the use of different types of online health apps available to the general population.

Prior research has revealed that internet attitude directly affects material access, the development of internet skills, and internet use [9]. Material access has significant relationships with both internet skills and internet use. Individuals with desktop computers, laptops, tablets, smartphones, and smart devices connect to the internet everywhere and at all times of the day and get more opportunities to develop varied skills and usage opportunities [7]. Internet skills affect the types of activities performed online and play a crucial role in translating uses into actual outcomes [10]. All stages have their own grounds of determination, interact to shape cumulative digital inequalities, and directly affect tangible health outcomes [9,10]. We therefore hypothesize that:

- Hypothesis 1 (H1): Internet attitude is positively associated with (1) material access, (2) internet skills, (3) health-related internet use, and (4) health outcomes.
- H2: Material access is positively associated with (1) internet skills, (2) health-related internet use, and (3) health outcomes.
- H3: Internet skills are positively associated with (1) health-related internet use and (2) health outcomes.
- H4: Health-related internet use is positively associated with health outcomes.

Health Statuses as Predictors of Internet Access

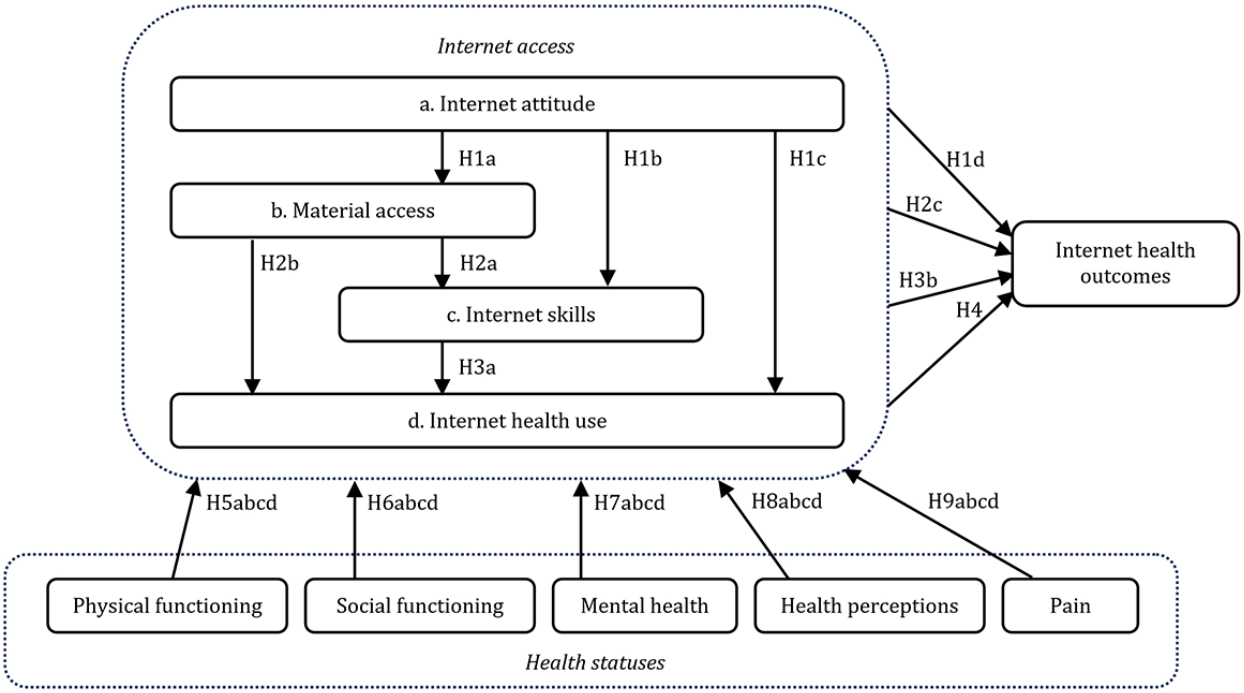
For the second goal of this paper, we focus on a range of health statuses pertaining to general functioning and well-being among the general population [11]. We first consider physical functioning, or the extent to which health interferes with a variety of functioning activities, such as participating in sports, carrying groceries, climbing stairs, or walking. Second, we consider social functioning, or the extent to which health interferes with normal social functioning activities, such as visiting friends. Third, mental health concerns one's general mood, including depression, anxiety, and psychological well-being. Fourth, health perception involves one's overall rating of current personal health. Finally, we consider the extent of bodily pain. We expect that these health statuses affect the different stages of internet access, as several studies have shown high rates of health-related internet use among those with medical conditions [12]. However, evidence on this relationship is inconclusive [12], as other studies have revealed that people who self-report being in good health are more likely to use the internet for health information [13] and that poor health inhibits particular stages of internet access [1,14]. For the different health statuses, causation could be argued in both directions, as compromised health might result in health-related internet use to become informed about specific conditions but might also restrict, for example, the use of certain devices or the development of internet skills. As the main purpose of this study is to assess the relationship between health statuses and internet access, we pose the following nondirectional hypotheses:

- H5: Physical functioning is associated with (1) internet attitude, (2) material access, (3) internet skills, and (4) health-related internet use.
- H6: Social functioning is associated with (1) internet attitude, (2) material access, (3) internet skills, and (4) health-related internet use.
- H7: Mental health is associated with (1) internet attitude, (2) material access, (3) internet skills, and (4) health-related internet use.
- H8: Health perceptions are associated with (1) internet attitude, (2) material access, (3) internet skills, and (4) health-related internet use.
- H9: Pain is associated with (1) internet attitude, (2) material access, (3) internet skills, and (4) health-related internet use.

Research Model

Figure 1 illustrates the research model built on the hypotheses. The model reflects resources and appropriation theory [6] by showing the core of the theory (the 4 phases of internet access) and considering personal categorical inequalities (in this contribution, the 5 health statuses). The internet health outcomes block reflects the potential benefits or outcomes from the 4 internet access phases.

Figure 1. Conceptual model and hypotheses.



Moderating Effects of Age and Education

The conceptual model in Figure 1 shows that the different health statuses are expected to support or inhibit internet access and, as such, obtain positive health outcomes. In this study, we further focus on the moderating roles of age and education, which represent important and common types of segmentation in digital inequality research [5]. We study to what extent age and educational differences exist in the contributions of the 5 health statuses to internet access. We expect contributions to become stronger for users over 65 years of age (seniors) when compared with the overall population and for less-educated users when compared with those with higher levels of educational attainment. Elderly and less educated individuals are more likely to perceive and actually suffer from limited health status [15]. Further examination of the moderating effects of age and education on internet access is important to explain differences in internet health outcomes. This approach further supports the development of health information and services aimed at different age and educational groups and future planning of the health care system for these specific groups.

Methods

Recruitment

This study used online surveys and drew upon a sample collected in the Netherlands. To obtain a representative sample of the population, we used PanelClix, a professional organization for market research. Members of the panel receive a small incentive for every survey they complete. In the Netherlands, 98% of the population uses the internet, closely representing the general population in terms of sociodemographic composition. We aimed to obtain a data set with approximately 1700 respondents over the age of 18 years. Eventually, this resulted in the collection of 1730 responses in a 2-wave study, both conducted over a 1-week period. The survey in the first wave (April 2020; n=2227) was specifically designed to gather background variables, including the different health statuses that are the topic of interest in this contribution. The survey furthermore included questions related to COVID-19. The average time required to complete this survey was 15-20 minutes. The survey in the second wave (November 2020; n=1730, 77.7%) was administered among respondents of the first wave and involved questions around the different internet

access stages, including internet motivation, material access, internet skills, internet health use, and health outcomes. The reason for administering a second survey among respondents of the first wave was a practical one: as the background variables were already collected, more space was available for questions related to internet access. Of the respondents of the first survey, 1730 (77.7%) completed the second survey. The average time required to complete this second survey was 20 minutes. During the first wave, 3 amendments to the sampling frame were made to ensure the representativeness of the Dutch population. Accordingly, the analyses revealed that respondents' gender, age, and formal education largely matched official census data. This was also the case for the sample that resulted from the second wave. See Table 1 for an overview.

Both online surveys followed Mahon's [16] recommendation to set an information sheet as the first page of the online survey in which potential respondents are required to check a box to indicate consent before accessing the survey. The survey used software that checked for missing responses and prompted users to respond. Both surveys were pilot-tested with 10 internet users over 2 rounds. Amendments were made based on the provided feedback. No major comments were provided in the second round.

Table 1. Characteristics of the study sample (N=1730).

Characteristics	Participants, n (%)
Gender	
Male	871 (50.3)
Female	859 (49.7)
Age (years; mean 50.24, SD 17.02)	
18-34	397 (22.9)
35-49	412 (23.8)
50-64	502 (29.0)
≥65	419 (24.2)
Educational level	
No diploma, primary or lower secondary diploma	516 (29.8)
Secondary diploma	602 (34.8)
Higher diploma	612 (35.4)

Ethical Considerations

To comply with requirements on privacy, collected data were anonymized by stripping IP addresses from the data set before the data files were saved to the researcher's computer.

Measures

Internet attitude was measured by 3 items adapted from the Digital Motivation Scale [17]. To measure *material access*, we considered a total of 7 different devices used to connect to the internet: desktop, laptop, tablet, smartphone, smart TV, game console, and smart device (eg, activity tracker; mean 3.43, SD 1.53). *Internet skills* were measured by the conceptual idea behind the Internet Skills Scale [9]. A 20-item measure was constructed in which items were scored on a 5-point scale. For health-related internet use, we used 6 items in which respondents were asked to indicate to what extent they used the internet for a particular online health activity. A 6-point scale was applied as an ordinal-level measure. Principal component analysis with varimax rotation was used to determine whether the items covered more underlying clusters, which was not the case. All items were retained in a single factor with an eigenvalue over 1.0, together accounting for 59% of the total variance. For *health outcomes*, we used 4 items that represent one's satisfaction with health-related achievements. All constructs exhibited high internal consistency; see Table 2.

The measures for the 5 considered health statuses were adapted from the Dutch version of the Medical Outcomes Study (MOS) Short-Form General Health Survey (SF-20) [18]. This instrument enables respondents to assess their general health and generates composite summary scores representing different health status. With the exception of physical pain, we normalized the scales, with higher scores representing better functioning, for *physical functioning*, *social functioning*, *mental health*, *health perception*, and *physical pain* (Table 3).

Gender was included as a dichotomous variable, and age was directly asked. Data on education were collected by degree. These were subsequently divided into 2 groups of low (ie, no diploma or primary or [lower] secondary education diploma) and high (ie, college and university) educational level attained.

Table 2. Items, descriptive statistics, and internal consistency (Cronbach α) for internet attitude, internet skills, health-related internet use, and health-related internet outcomes.

Items	Mean (SD)
Internet attitude (α=.74)[a], mean 4.10, SD 0.70	
Technologies, such as the internet and mobile phones, make life easier.	4.29 (0.83)
I feel that people pressure me to be constantly connected (recoded).	4.03 (1.23)
There are many things on the internet that are good for people like me.	3.89 (0.85)
Internet skills (α=.96)[b], mean 3.45, SD 0.96	
I know how to upload files.	3.16 (1.07)
I know how to adjust privacy settings.	3.54 (1.04)
I know how to use my smartphone as a hotspot.	4.11 (1.55)
I know how to check whether the information I find online is true.	3.33 (1.21)
I find it easy to decide what the best keywords are.	4.17 (1.02)
I know how to figure out whether a website can be trusted.	3.71 (1.22)
I know how to store photos, documents, or other files in the cloud (eg, Google Drive, iCloud).	3.76 (1.41)
I know how to keep track of the costs of mobile app use.	4.17 (1.41)
I know how to change with whom I share content (eg, friends, friends of friends, or the public).	4.28 (1.13)
I know how to block messages from people I do not want to have anything to do with anymore.	4.16 (1.14)
I know what pictures of me or others I can share online.	4.23 (1.11)
I know how to turn off my location on a smartphone.	3.24 (1.19)
I know how to reach people with my digital creations.	3.66 (1.37)
I know how to create videos or selfies to which others will react positively.	4.18 (1.32)
I know how to create digital materials to express my ideas.	3.60 (1.35)
I know how to block unwanted popup messages or ads.	3.59 (1.38)
I know how to post homemade videos or music online.	3.64 (1.46)
I know how to make basic changes to the content that others have produced.	3.48 (1.38)
I know which (copy) rights apply to online material.	3.57 (1.29)
I know how to increase the number of followers of my profile on social media.	3.45 (2.05)
Health-related internet use (α=.86)[c], mean 2.08, SD 0.86	
Finding information about your health or medical care	2.60 (1.02)
Contacting a physician or medical specialist	1.94 (0.98)
Talking to others about your personal health	1.91 (1.23)
Participating in an online training or health program	1.76 (1.18)
Finding information or watching videos about improving your fitness/health	2.05 (1.27)
Using an app to check your health status or treatment	1.96 (1.33)
Health-related internet outcomes (α=.85)[d], mean 2.13, SD 1.39	
The way the last advice, program, or app you used affected your health	2.04 (1.56)
The feeling about your fitness/health that online information gives you	2.23 (1.59)
The latest online health information or online advice that you applied	3.03 (2.03)
The way you have adapted your behavior based on online health information	2.11 (1.53)

[a]A 5-point agreement scale ranging from “strongly disagree” to “strongly agree.”

[b]A 5-point truth scale ranging from “not at all true of me” to “very true of me.”

[c]A 6-point frequency scale ranging from “never” to “multiple times a day.”

[d]A 5-point satisfaction scale ranging from “very dissatisfied” to “very satisfied.”

Table 3. Items, descriptive statistics, and internal consistency (Cronbach α) for health state variables.

Items	Mean (SD)
Physical functioning (α=.89)[a], mean 1.75, SD 0.34	
Vigorous activities, such as lifting heavy objects, running, or participating in strenuous sports	1.57 (0.50)
Moderate activities, such as moving a table or carrying groceries	1.77 (0.42)
Walking uphill or climbing a few steps without resting	1.74 (0.44)
Bending or lifting or stooping	1.73 (0.44)
Walking 1 block	1.83 (0.38)
Eating, dressing, bathing, or using the toilet	1.89 (0.31)
Social functioning[b]	
My health regularly limits me in social activities (eg, visiting friends or family)—recoded.	3.81 (1.16)
Mental health (α=.85)[b], mean 3.65, SD 0.77	
I regularly feel depressed and gloomy (recoded).	3.43 (1.05)
I am often so sad that nothing can cheer me up (recoded).	3.60 (0.87)
I am regularly nervous (recoded).	3.65 (1.10)
I usually feel calm and composed.	3.66 (0.84)
I feel happy most of the time.	4.05 (1.01)
Health perception (α=.86)[b], mean 3.39, SD 0.85	
I am a little sick (recoded).	3.72 (1.18)
I am as healthy as anyone I know.	3.22 (1.04)
My health is excellent.	3.28 (1.06)
I have been feeling bad lately (recoded).	3.73 (1.05)
Physical pain[c]	
Have you experienced any physical pain in the past 4 weeks?	3.67 (1.26)

[a]Did your health condition limit you in any of the following activities last year? If so, for how long? Yes, longer than 3 months/Yes, less than 3 months/No → transposed to No (1)/Yes (2).

[b]A 5-point agreement scale ranging from “strongly disagree” to “strongly agree.”

[c]A 5-point scale ranging from “heavy pain” to “no pain.”

Statistical Analysis

To test the first hypothesized relationships, we applied path analysis with Amos 23 (IBM Corporation). To obtain a comprehensive model fit, we included the suggested indices by Hair et al [19]: the χ^2 statistic, the ratio of χ^2 to its *df*, the standardized root mean residual (SRMR<0.08), the Tucker-Lewis index (TLI>0.90), the comparative fit index (CFI>0.95), and the root mean square error of approximation (RMSEA<0.06). These fit indices are typically used to represent the 3 categories of model fit: absolute, parsimonious, and incremental. We added covariates between the health status variables. The correlations between internet attitude, material access, internet skills, internet health use, and health outcomes were not high enough to cause multicollinearity concerns. To test for moderator effects of age and education, we applied multigroup analyses. First, the model was estimated for each of the subgroups separately to confirm its acceptable fit for each group. Then, multigroup analysis was used to test the significance of the χ^2 difference.

Results

Measurement Model and Hypotheses

To test the hypothesized relationships, we started by examining the basic assumptions of path analysis. Normality, kurtosis, and skewness did not differ significantly from acceptable criteria, and there were no outliers or multicollinearity beyond what would theoretically be expected. The structural model with coefficients and variances explained is presented in Figure 2. The results of the fit statistics indicated a good model fit: χ^2_5=23.68; χ^2/df=4.74; SRMR=0.01; TLI=0.96; CFI=1.00, RMSEA=0.05 (90% CI 0.03-0.07). The magnitudes and significance of the direct, indirect, and total path coefficients are shown in Table 4. The significance of the indirect effects was examined using bootstrapping procedures [20] and the Monte Carlo method for assessing mediation [21,22].

The first hypotheses concerning the internet access stages and outcomes (H1-H4) are supported, with the exception of H1c. Internet attitude has a negative direct path to internet health use,

and the total effect is 0. For the other hypotheses, all direct and indirect paths are positive and significant. See Table 4.

For the second set of hypotheses (concerning the health statuses), first Table 4 shows that physical functioning is directly and indirectly related to all 4 internet access stages (supporting H5a-d). Physical functioning contributes positively to internet attitude, material access, and internet skills but negatively to internet health use. Second, social functioning is directly and indirectly related to internet attitude, internet skills, and internet health use (supporting H6a,c,d). Social functioning contributes negatively to internet attitude and positively to internet skills and internet health use. There is a small indirect negative contribution to material access (partly supporting H6b). Third, the results revealed that mental health contributes positively to internet attitude and negatively to material access and internet health use (supporting H7a,b,d). There is no significant direct or indirect contribution to internet skills (rejecting H7c). Fourth, perceived health has a direct positive contribution to material access, internet skills, and internet health use (supporting H8b-d). There is no significant contribution to internet attitude (rejecting H8a). Finally, physical pain contributes positively to internet attitude and material access (supporting H8a,b). There are positive indirect contributions to internet skills and internet health use (partly supporting H8c,d).

Figure 2. Structural model with path coefficients. Note: Path coefficients are significant at $P<.05$. Squared multiple correlations are underlined. ns: not significant.

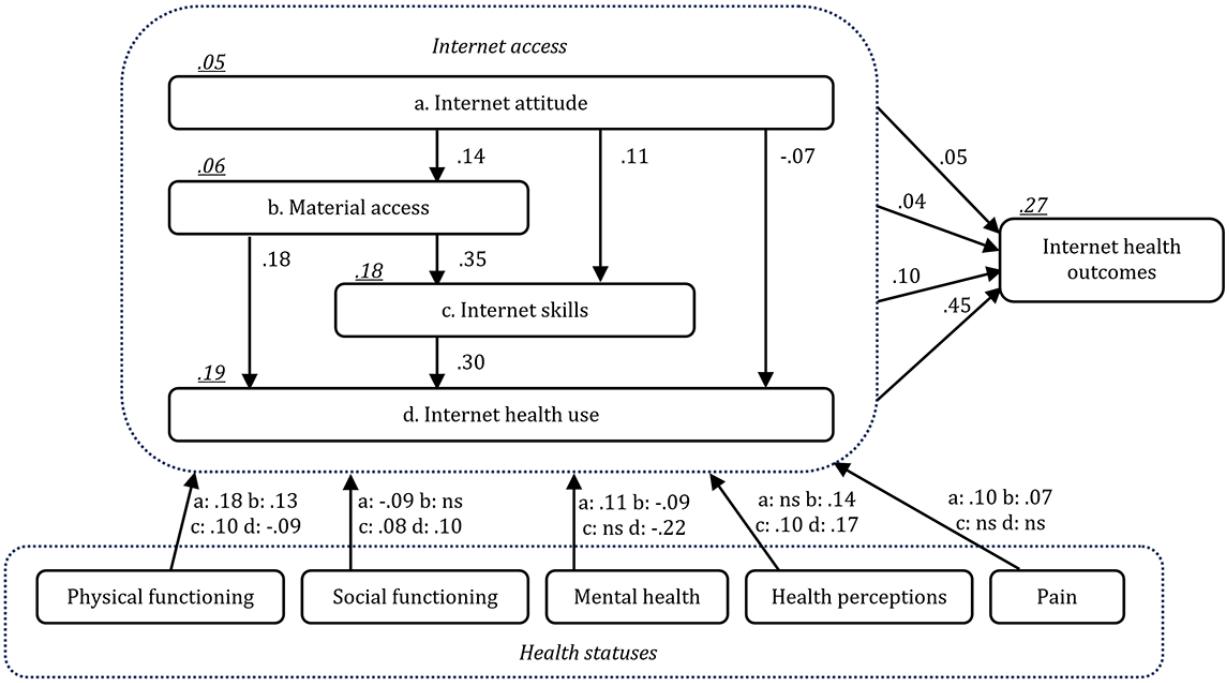


Table 4. Significant direct, indirect, and total effects (standardizes regression weights and significance).

Path	Direct effects		Indirect effects		Total effects	
	β	*P* value	β	*P* value	β	*P* value
Internet attitude → health outcome	.05	.01	.03	.01	.08	.01
Material access → health outcome	.04	.04	.16	.01	.20	.01
Internet skills → health outcome	.10	.01	.13	.03	.23	.01
Internet health use → health outcome	.45	.02	N/A[a]	N/A	.45	.02
Internet attitude → material access	.14	.02	N/A	N/A	.14	.02
Internet attitude → digital skills	.11	.01	.05	.01	.16	.01
Internet attitude → internet health use	–.07	.03	.07	.01	.00	.50
Material access → internet skills	.35	.02	N/A	N/A	.35	.01
Material access → internet health use	.18	.01	.10	.01	.28	.01
Internet skills → internet health use	.30	.04	N/A	N/A	.30	.01
Physical functioning → internet attitude	.18	.02	N/A	N/A	.18	.02
Physical functioning → material access	.13	.01	.03	.01	.16	.02
Physical functioning → internet skills	.10	.01	.08	.01	.18	.01
Physical functioning → internet health use	–.09	.01	.07	.01	–.02	.62
Physical functioning → health outcomes	N/A	N/A	.03	.12	.03	.12
Social functioning → internet attitude	–.09	.01	N/A	N/A	–.09	.01
Social functioning → material access	.02	.49	–.01	.01	.01	.77
Social functioning → internet skills	.08	.03	–.01	.66	.07	.09
Social functioning → internet health use	.10	.01	.03	.07	.13	.01
Social functioning → health outcomes	N/A	N/A	.06	.02	.06	.02
Mental health → internet attitude	.11	.02	N/A	N/A	.11	.02
Mental health → material access	–.09	.02	.02	.01	–.07	.02
Mental health → internet skills	.01	.80	–.02	.26	–.01	.74
Mental health → internet health use	–.22	.01	–.02	.05	–.24	.01
Mental health → health outcomes	N/A	N/A	–.11	.02	–.11	.02
Perceived health → internet attitude	–.04	.36	N/A	N/A	–.04	.36
Perceived health → material access	.14	.02	–.01	.31	.13	.02
Perceived health → internet skills	.10	.03	.04	.01	.14	.02
Perceived health → internet health use	.17	.01	.07	.02	.24	.02
Perceived health → health outcomes	N/A	N/A	.12	.02	.12	.02
Physical pain → internet attitude	–.10	.01	N/A	N/A	.10	.01
Physical pain → material access	–.07	.01	–.01	.02	–.08	.01
Physical pain → internet skills	–.01	.65	–.04	.01	–.05	.09
Physical pain → internet health use	–.02	.65	–.03	.03	–.05	.19
Physical pain → health outcomes	N/A	N/A	–.04	.02	–.04	.02

[a] N/A: not applicable.

Moderator Effects

We tested for the significance of the χ^2 difference between 2 specified age groups (<65 and ≥65 years) and between 2 educational groups (low and high). The results showed that for both age (χ^2/df=3.716, *P*<.001, TLI=0.946, CFI=0.994, RMSEA=0.04 [95% CI 0.03-0.05]) and education (χ^2/df=2.944, *P*<.001, TLI=0.962, CFI=0.996, RMSEA=0.03 [95% CI 0.02-0.05]), there are moderation effects on the overall model χ^2. Table 5 shows the results of the direct path coefficient

comparison between the 2 age groups and between the 2 educational groups.

Concerning age and internet access, Table 5 shows that the direct path coefficients from internet attitude to material access and internet skills are significantly larger for seniors. Furthermore, internet attitude contributes negatively to internet health use and positively to health outcomes in the group aged below 65 years. The contribution of material access to internet skills is slightly larger in the senior group, and the contribution of internet skills to internet health use is slightly smaller. In terms of age and the different health statuses, Table 5 reveals that physical functioning contributes positively to internet attitude, material access, and internet skills and negatively to internet health use in the group aged under 65 years. In the oldest age group, physical functioning contributes only positively to internet attitude. Social functioning contributes negatively to internet attitude and positively to internet skills and internet health use in the group aged under 65 years, while there are no direct significant contributions in senior group. In the group aged under 65 years, mental health contributes positively to internet attitude and negatively to material access and internet health use. In the senior group, there are positive contributions to material access and internet skills. The negative contribution of mental health to internet health use is significantly larger in the younger group. In this group, perceived health contributes negatively to internet attitude and positively to material access, internet skills, and internet health use. In the senior group, there is a negative contribution to material access. Finally, physical pain contributes negatively to internet attitude in both age groups, to material access in the younger group, and to internet health use in the senior group.

For education and internet access, Table 5 shows that the magnitude of the contribution of internet attitude to material access is larger among the less educated. Furthermore, internet attitude contributes positively to internet skills in this group. The contribution of material access to internet skills is also larger in the lower-educated group, while the contribution to internet health use is larger in the higher-educated group. The contribution of internet skills to internet health use is larger in the less-educated group. In relation to the different health statuses, the results showed that physical functioning contributes significantly more to internet attitude in the lower-educated group, while the positive contribution to material access is larger in the higher-educated group. For social functioning, there is a negative effect on internet attitude and a positive effect on material access in the higher-educated group. In the lower-educated group, there are positive effects on internet skills and internet heath use. Concerning mental health, the positive contribution to internet attitude and the negative contribution to material access are larger in the higher-educated group. For perceived health, in the higher-educated group, there is a significant effect on material access. Furthermore, there is a larger significant effect on internet health use in the higher-educated group. Finally, in the higher-educated group, there is a negative effect of physical pain on material access.

Table 5. Direct path coefficient comparisons for age and education.

Path	Age<65 years		Age≥65 years		Low education level		High education level	
	β	*P* Value	β	*P* Value	β	*P* Value	β	*P* Value
Internet attitude → health outcome	.06	.01	.03	.46	.03	.27	.08	.02
Material access → health outcome	.04	.18	–.00	.93	.05	.06	–.01	.74
Internet skills → health outcome	.04	.12	.11	.02	.07	.01	.12	<.001
Internet health use → health outcome	.46	<.001	.42	<.001	.45	<.001	.48	<.001
Internet attitude → material access	.13	<.001	.24	<.001	.21	<.001	.13	.002
Internet attitude → internet skills	.11	<.001	.14	.003	.14	<.001	.01	.77
Internet attitude → internet health use	–.08	.003	.08	.10	–.05	.06	–.07	.08
Material access → internet skills	.28	<.001	.31	<.001	.36	<.001	.29	<.001
Material access → internet health use	.16	<.001	.15	.002	.16	<.001	.21	<.001
Internet skills → internet health use	.26	<.001	.23	<.001	.30	<.001	.25	<.001
Physical functioning → internet attitude	.18	<.001	.19	.01	.21	<.001	.10	.04
Physical functioning → material access	.08	.03	.13	.08	.10	.01	.16	<.001
Physical functioning → internet skills	.08	.01	.02	.80	.10	.01	.10	.04
Physical functioning → internet health use	–.10	.003	–.11	.10	–.08	.04	–.11	.02
Social functioning → internet attitude	–.13	<.001	.03	.68	–.00	.94	–.25	<.001
Social functioning → internet access	.05	.21	–.08	.24	–.06	.15	.19	<.001
Social functioning → internet skills	.07	.04	.13	.07	.10	.01	.06	.31
Social functioning → internet health use	.11	.002	.01	.91	.10	.01	.08	.12
Mental health → internet attitude	.17	<.001	–.03	.64	.10	.01	.14	.003
Mental health → material access	–.10	.002	.14	.01	–.08	.02	–.12	.01
Mental health → internet skills	.01	.71	.12	.02	–.03	.33	.09	.06
Mental health → internet health use	–.21	<.001	–.16	.004	–.21	<.001	–.22	<.001
Perceived health → internet attitude	–.10	.03	.08	.31	–.04	.45	–.06	.34
Perceived health → material access	.19	<.001	–.11	.02	.08	.10	.24	<.001
Perceived health → internet skills	.12	.01	.08	.26	.11	.01	.09	.16
Perceived health → internet health use	.18	<.001	.10	.20	.15	<.001	.20	<.001
Physical pain → internet attitude	–.09	.01	–.12	.04	–.10	.01	–.11	.02
Physical pain → material access	–.09	.01	–.03	.59	–.04	.22	–.13	.01
Physical pain → internet skills	–.01	.71	–.00	.94	–.01	.78	–.02	.63
Physical pain → internet health use	–.00	.99	–.11	.05	–.01	.78	–.04	.44

Discussion

Principal Findings

This paper aimed to provide a comprehensive view of digital inequality in relation to different health statuses among the Dutch population. The study's first goal was to reveal to what extent the process of internet access is important to obtain health outcomes. Internet attitude increases the likelihood of improving material access, the development of internet skills, and internet health use, suggesting that making online health apps attractive for larger segments of the population is an important objective. Material access, considered in this study as the diversity of the devices used, is highly relevant, as it has significant relationships with internet skills and internet health use. Individuals with different devices to connect to the internet everywhere and at all times of the day have more opportunities to develop internet skills and use online health apps. Internet skills are, in turn, required to use online health apps. The sequential nature of the access stages does not suggest that improving material access will automatically result in better internet skills or that a high level of internet skills will automatically result in a large variety of health-related internet use; all stages are, however, necessary conditions. The results furthermore revealed that all 4 access stages directly contribute to obtaining positive health outcomes, suggesting that to make online health care attainable for the general population, interventions should focus *simultaneously* on all stages. For example, attitudes might be improved by

considering issues of accessibility and usability of online health information and services, material access by offering schemes such as device donation, internet skills by training programs tailored to the needs of people with different health statuses, and online health apps by awareness programs. Such approaches would require government, public, private, and nonprofit sector organizations to collaborate.

The second goal of this paper was to reveal to what extent different health statuses among the general population relate to the internet access stages and thus to internet health outcomes. The results confirmed that digital inequality research would benefit from considering health as a predictor of internet attitude, material access, internet skills, internet health use, and health outcomes. However, a general conclusion is that we should go beyond single self-reported measures of health, as different health statuses among the general population make unique contributions to the different internet access stages:

- Physical functioning contributes to internet attitude, material access, and internet skills, likely because physical limitations impact the process of taking up or learning how to use technologies (eg, in the case of smaller tablets or smartphones) [23]. Those with better physical functioning make less use of online health information and services as they have a relatively low need. Similarly, people with specific diseases that hinder physical functioning have less information need about their disease if they experience less limitations (eg, in the case of rheumatoid arthritis) [24].
- Better social functioning contributes to better material access and higher levels of internet skills. The importance of social bonds to use technology has long been established [25], and support from family, friends, or those that are important to the individual's life contributes to learning to use a device or improving internet skills [26]. This is further strengthened when mobile phones, tablets, or laptops further enhance social connections and communication. Note that for internet skills, research has shown that informal support mainly works to apply basic skills [27]. The use of online health information and services is higher for those with poorer social functioning. This suggests that those whose health restricts people from visiting friends and family are more likely to seek health information online. This might be the result of a higher need for online health information and services but also of online health information serving as a substitute for information received from peers.
- Concerning mental health, the results revealed a positive contribution to internet attitude but a negative contribution to material access. An explanation might be that those suffering from mental health issues are more likely to experience excessive internet use, which is supported by the use of multiple devices to provide instant access at all times [28]. Furthermore, mental health negatively contributes to internet health use. As mental health is reflective of general distress, it causes people to turn to the internet for health information and services [29], apparently despite their less positive attitude toward the internet.
- People who perceive their health as higher have greater levels of material access and internet skills. A possible explanation might be that higher health perceptions foster social interactions that are supported by material access and higher levels of internet skills in the case of online social networking. The higher use of online health information and services among those with higher health perceptions seems to be inconsistent with prior research [30]. This discrepancy might be related to the influence of the COVID-19 pandemic in the survey period.
- Like poor physical functioning, physical pain negatively affects internet attitude and material access, suggesting that physical pain limits the use of certain devices and the process of learning how to use the internet.

In relation to our third goal, the general conclusion is that the contributions of the health statuses to the internet access stages differ for age and education. The main findings concerning age are that for seniors:

- internet attitude plays a more important role in obtaining material access than for those aged under 65 years. An important reason for seniors not to go online is a less favorable attitude toward the internet [31]. A positive, guided experience with the internet might motivate seniors to move to the following stages of internet access [31]. Furthermore, seniors are most likely to benefit most from accessible and usable apps [32].
- mental health plays a larger role in obtaining material access and developing internet skills. This suggests that seniors with mental health issues have a relatively high need for support, a worthwhile finding as online health interventions can reduce their mental health problems [33].
- perceived poor health hinders material access, suggesting that seniors who believe they are in poor health consider this as a barrier to interact with computer devices. This is a missed opportunity, as smartphones, tablets, or laptops might also be used as tools to enhance their perceived health [34].

The main findings concerning education are that for those with *lower* levels of education:

- internet attitude plays a larger role in obtaining material access, consistent with prior research that showed that education positively affects internet attitude [9]. Similar suggestions discussed for seniors apply, although specific approaches will be required.
- physical functioning is relatively important for developing a favorable internet attitude. This might be explained by the fact that lower-educated individuals are more likely to suffer from limitations in physical functioning [35], which could hinder the process of taking up and learning how to use the internet.
- social functioning plays a relatively important role in the development of internet skills and the use of online health information and services. Unfortunately, lower-educated individuals are less likely to perceive higher levels of support in relation to health [36], making organizing access to support an important objective.
- perceived health is relatively important for the development of internet skills. This suggests that lower-educated people who believe they are in poor health are more in need for

skills training to make use of online health information and services as compared to their higher-educated counterparts.

Limitations

A few limitations should be noted. The first is the study's cross-sectional design, which did not allow confirmation of causal inferences about the association between health statuses and internet access. Furthermore, we focused on the general population, and the baseline status of the different health statuses varied slightly. Effects might have been stronger when targeting more people with serious conditions in relation to the 5 health statuses, although that was not the purpose of this study. Finally, we encourage further qualitative research to focus on the barriers and facilitators for people with different health statuses when using the internet to support their health needs.

Conclusion

To obtain positive health outcomes and make online health care attainable for the general population, interventions should focus simultaneously on internet attitude, material access, internet skills, and internet health apps. However, issues of equality need to be considered and digital inequality research would benefit from considering health as a predictor of all 4 internet access stages and health outcomes. Furthermore, studies among the general population should go beyond single self-reported measures of health as physical functioning, social functioning, mental health, perceived health, and physical pain all demonstrated unique contributions to the internet access stages. The general conclusion is that different health statuses affect internet access stages in different ways and, consequently, the health-related opportunities that the internet offers. Further complicating this issue is that such influence is moderated by age and education.

Conflicts of Interest

None declared.

References

1. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
2. Shim M, Jo HS. What quality factors matter in enhancing the perceived benefits of online health information sites? Application of the updated DeLone and McLean Information Systems Success Model. Int J Med Inform 2020 May;137:104093. [doi: 10.1016/j.ijmedinf.2020.104093]
3. Atanasova S, Kamin T, Petrič G. Exploring the benefits and challenges of health professionals' participation in online health communities: emergence of (dis)empowerment processes and outcomes. Int J Med Inform 2017 Feb;98:13-21. [doi: 10.1016/j.ijmedinf.2016.11.005] [Medline: 28034408]
4. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St Martin's Press; 2018.
5. Scheerder A, van Deursen AJAM, van Dijk JAGM. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telem Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
6. Van Dijk JAGM. The Deepening Divide: Inequality in the Information Society. London, UK: Sage; 2005.
7. van Deursen AJAM, van Dijk JAGM. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb 07;21(2):354-375 [FREE Full text] [doi: 10.1177/1461444818797082] [Medline: 30886536]
8. van Deursen AJAM, Helsper EJ, Eynon R. Development and validation of the Internet Skills Scale (ISS). Inf Commun Soc 2015 Aug 25;19(6):804-823. [doi: 10.1080/1369118x.2015.1078834]
9. van Deursen AJAM, van Dijk JAGM. Toward a multifaceted model of internet access for understanding digital divides: an empirical investigation. Inf Soc 2015 Sep 11;31(5):379-391. [doi: 10.1080/01972243.2015.1069770]
10. Van Deursen AJAM, Helsper EJ. Collateral benefits of internet use: explaining the diverse outcomes of engaging with the Internet. New Media Soc 2018 Jul 30;20(7):2333-2351 [FREE Full text] [doi: 10.1177/1461444817715282] [Medline: 30581362]
11. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. Med Care 1988 Jul;26(7):724-735. [doi: 10.1097/00005650-198807000-00007] [Medline: 3393032]
12. Mariano J, Marques S, Ramos MR, de Vries H. Internet use by middle-aged and older adults: longitudinal relationships with functional ability, social support, and self-perceptions of aging. Psychol Aging 2021 Dec;36(8):983-995. [doi: 10.1037/pag0000643] [Medline: 34498895]
13. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
14. Goldner M. How health status impacts the types of information consumers seek online. Inf Commun Soc 2006 Dec;9(6):693-713. [doi: 10.1080/13691180601063982]
15. Barrett AE. Socioeconomic status and age identity: the role of dimensions of health in the subjective construction of age. J Gerontol B Psychol Sci Soc Sci 2003 Mar 01;58(2):S101-S109. [doi: 10.1093/geronb/58.2.s101] [Medline: 12646599]

16. Mahon PY. Internet research and ethics: transformative issues in nursing education research. J Prof Nurs 2014 Mar;30(2):124-129. [doi: 10.1016/j.profnurs.2013.06.007]
17. Helsper EJ, Smirnova S, Robinson D. Digital Reach Survey. 2017. URL: http://www.lse.ac.uk/media-and-communications/research/research-projects/disto/disto-youth [accessed 2022-10-12]
18. Kempen G, Brilman E, Heyink J, Ormel J. MOS Short-Form General Health Survey. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken (NCG); 1995.
19. Hair J. Multivariate Data Analysis. Upper Saddle River, NJ: Pearson Prentice Hall; 2006.
20. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/BRM.40.3.879]
21. Mackinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivariate Behav Res 2004 Jan 01;39(1):99-128 [FREE Full text] [doi: 10.1207/s15327906mbr3901_4] [Medline: 20157642]
22. Selig JP, Preacher KJ. Monte Carlo Method for Assessing Mediation: An Interactive Tool for Creating Confidence Intervals for Indirect Effects. 2008. URL: http://www.quantpsy.org [accessed 2022-10-12]
23. Olphert W, Damodaran L. Older people and digital disengagement: a fourth digital divide? Gerontology 2013;59(6):564-570. [doi: 10.1159/000353630] [Medline: 23969758]
24. Meesters J, de Boer I, van den Berg M, Fiocco M, Vliet Vlieland T. Unmet information needs about the delivery of rheumatology health care services: a survey among patients with rheumatoid arthritis. Patient Educ Couns 2011 Nov;85(2):299-303 [FREE Full text] [doi: 10.1016/j.pec.2010.10.004] [Medline: 21051173]
25. Fulk J, Schmitz J, Steinfield CW. A social influence model of technology use. In: Organizations and Communication Technology. New York, NY: Sage; 1990:117-140.
26. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice-Hall; 1977.
27. van Deursen AJAM, Courtois C, van Dijk JAGM. Internet skills, sources of support, and benefiting from internet use. Int J Hum-Comput Interact 2014 Mar 07;30(4):278-290. [doi: 10.1080/10447318.2013.858458]
28. Yang C, Choe B, Baity M, Lee J, Cho J. SCL-90-R and 16PF profiles of senior high school students with excessive internet use. Can J Psychiatry 2005 Jun 01;50(7):407-414. [doi: 10.1177/070674370505000704] [Medline: 16086538]
29. Bowen D. Predictors of women's Internet access and Internet health seeking. Health Care Women Int 2003 Dec;24(10):940-951. [doi: 10.1080/07399330390244130] [Medline: 14742131]
30. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
31. van Deursen AJAM, Helsper EJ. A nuanced understanding of internet use and non-use among the elderly. Eur J Commun 2015 Apr 09;30(2):171-187. [doi: 10.1177/0267323115578059]
32. Lee B, Chen Y, Hewitt L. Age differences in constraints encountered by seniors in their use of computers and the internet. Comput Hum Behav 2011 May;27(3):1231-1237. [doi: 10.1016/j.chb.2011.01.003]
33. Jones SL, Hadjistavropoulos HD, Soucy JN. A randomized controlled trial of guided internet-delivered cognitive behaviour therapy for older adults with generalized anxiety. J Anxiety Disord 2016 Jan;37:1-9. [doi: 10.1016/j.janxdis.2015.10.006] [Medline: 26561733]
34. Dasgupta D, Chaudhry B, Koh E, Chawla NV. A survey of tablet applications for promoting successful aging in older adults. IEEE Access 2016;4:9005-9017. [doi: 10.1109/ACCESS.2016.2632818]
35. Zajacova A, Montez JK. Physical functioning trends among US women and men age 45-64 by education level. Biodemography Soc Biol 2017 Mar 13;63(1):21-30 [FREE Full text] [doi: 10.1080/19485565.2016.1263150] [Medline: 28287310]
36. Croezen S, Picavet HSJ, Haveman-Nies A, Verschuren WM, de Groot LC, van't Veer P. Do positive or negative experiences of social support relate to current and future health? Results from the Doetinchem Cohort Study. BMC Public Health 2012 Jan 21;12(1):65 [FREE Full text] [doi: 10.1186/1471-2458-12-65] [Medline: 22264236]

Abbreviations

CFI: comparative fit index
RMSEA: root mean square error of approximation
SRMR: standardized root mean residual
TLI: Tucker-Lewis index

Edited by R Kukafka, G Eysenbach; submitted 09.03.22; peer-reviewed by C Jacob, L Welch; comments to author 28.06.22; revised version received 14.07.22; accepted 08.10.22; published 21.10.22.

Please cite as:
van Deursen AJAM
General Health Statuses as Indicators of Digital Inequality and the Moderating Effects of Age and Education: Cross-sectional Study
J Med Internet Res 2022;24(10):e37845
URL: https://www.jmir.org/2022/10/e37845
doi:10.2196/37845
PMID:36269664

Original Paper

The Effects of the Use of Patient-Accessible Electronic Health Record Portals on Cancer Survivors' Health Outcomes: Cross-sectional Survey Study

Piper Liping Liu[1], PhD; Xinshu Zhao[1], PhD; Jizhou Francis Ye[1], MA

Department of Communication, Faculty of Social Sciences, University of Macau, Taipa, Macao

Corresponding Author:
Xinshu Zhao, PhD
Department of Communication
Faculty of Social Sciences
University of Macau
Room 2051, Humanities & Social Sciences Building (E21B)
Taipa
Macao
Phone: 853 8822 8837
Email: xszhao@um.edu.mo

Abstract

Background: In the past decade, patient-accessible electronic health record (PAEHR) systems have emerged as an important tool for health management both at the hospital level and individual level. However, little is known about the effects of PAEHR portals on the survivorship of patients with chronic health conditions (eg, cancer).

Objective: This study aims to investigate the effects of the use of PAEHR portals on cancer survivors' health outcomes and to examine the mediation pathways through patient-centered communication (PCC) and health self-efficacy.

Methods: Data for this study were derived from the Health Information National Trends Survey (HINTS 5, Cycle 4) collected from February 2020 to June 2020. This study only involved respondents who reported having been diagnosed with cancer (N=626). Descriptive analyses were performed, and the mediation models were tested using Model 6 from the SPSS macro PROCESS. Statistically significant relationships among PAEHR portal use, PCC, health self-efficacy, and physical and psychological health were examined using bootstrapping procedures. In this study, we referred to the regression coefficients generated by min-max normalization as percentage coefficients (b_p). The 95% bootstrapped CIs were used with 10,000 resamplings.

Results: No positive direct associations between PAEHR portal use and cancer survivors' health outcomes were found. The results supported the indirect relationship between PAEHR portal use and cancer survivors' psychological health via (1) PCC (b_p=0.029; β=.023, 95% CI .009-.054), and (2) PCC and health self-efficacy in sequence (b_p=0.006; β=.005, 95% CI .002-.014). Besides, the indirect association between PAEHR portal use and cancer survivors' physical health (b_p=0.006; β=.004, 95% CI .002-.018) via sequential mediators of PCC and health self-efficacy was also statistically acknowledged.

Conclusions: This study offers empirical evidence about the significant role of PAEHR portals in delivering PCC, improving health self-efficacy, and ultimately contributing to cancer survivors' physical and psychological health.

(J Med Internet Res 2022;24(10):e39614) doi:10.2196/39614

KEYWORDS

electronic health record; patient-centered care; health self-efficacy; cancer survivors; physical health; psychological health

Introduction

Cancer is among the leading causes of death worldwide, accounting for about 10 million deaths in 2020 [1]. In 2021, 1.9 million new cancer cases were diagnosed and over 600,000 cancer deaths were estimated in the United States [2]. Due to the growing and aging population as well as increases in early diagnoses and advances in cancer treatments, the number of cancer survivors continues to increase [3]. According to the National Cancer Institute, "An individual is considered a cancer survivor from the time of diagnosis, through the balance of his or her life" [4]. Cancer is viewed as a chronic illness, and cancer survivors face ongoing health challenges that call for unique

and long-term survivorship care. This is because physical problems such as functional disability and impairment and psychological disorders due to illness and aggressive treatments might persist throughout cancer survivors' lifetime [3,5]. As such, delivering high-quality and long-term health care for cancer survivors becomes a major challenge facing public health.

The maintenance of long-term cancer treatment plans requires effective patient-provider communication and coordination of cancer survivorship care [6,7]. Health care information technology has brought about a massive change in cancer care. The transition to patient-accessible electronic health record (PAEHR) systems has changed the way patients and providers engage in health care by facilitating access to patient information (eg, test results) [8], allowing timely and efficient patient-provider communication [9], reducing medical errors [10], educating patients with accessible and affordable health materials [11], and enhancing the privacy and security of patient data [12]. Therefore, researchers generally agree that PAEHR portals have the potential to improve health through evidence-based medicine and effective care coordination [13]. For instance, Wani and Malhotra [14] provided empirical evidence supporting that the assimilation of PAEHRs at a hospital-wide level can help deliver quality care and services, which in turn improve patients' health outcomes. A systematic review conducted by Kruse et al [13] identified a variety of facilitators of PAEHRs that can improve population health, including the enhancement in productivity/efficiency, the increase in the quality of patient data, and more flexible data management. Nevertheless, the majority of existing studies have inevitably investigated the PAEHR system from perspectives on professionals' innovation adoption [15] or organizational management [16]. There remains a paucity in the literature on the use of PAEHR portals and health outcomes from patient perspectives. To address this literature gap, our study aims to investigate how PAEHR portal use influences cancer survivors' health outcomes.

The Chronic Care Model (CCM) provides a framework for understanding the mechanisms through which health care provided via PAEHR portals influences patients' health outcomes [17]. Six key interdependent components of CCM that are essential for care delivery have been identified: (1) health system support, (2) delivery system design, (3) clinical information systems, (4) community resources, (5) decision support, and (6) self-management support. Researchers suggest that the PAEHR portal may be a prominent tool that incorporates the key elements of CCM and determines the success of care delivery and health management [18]. CCM relies on the use of health information technology for both public and private health care systems to facilitate the provision of longitudinal and patient-centered care, improve patient engagement, and empower patients with self-care skills to manage chronic illness [18,19]. Gee et al [19] proposed a revised CCM—eHealth enhanced CCM (eCCM)—and explicated that the use of eHealth technologies can help improve chronic care (eg, through patient-centered communication [PCC], clinical decision support, information provision, health education). Consequently, experienced PAEHR users have higher health self-efficacy and can achieve improved health outcomes [19].

Proponents of the eCCM contend that eHealth adoption, referred to in this study as PAEHR portal use, is likely to impact health outcomes through indirect pathways, which comprise proximal outcomes (eg, effective patient-provider communication) of eHealth that then influence health or that contribute to intermediate outcomes (eg, health self-efficacy) that lead to improved distal health outcomes [19]. Rathert et al [20] provide tentative support for the serial mediation effect of PCC and health self-efficacy in the relationship between PAEHR portal use and health outcomes. PCC is about delivering health care that relies upon effective communication and empathy to meet individual patient preferences, needs, and values [21,22]. Health self-efficacy refers to people's beliefs regarding one's capabilities to execute the courses of action to improve health [23]. There is a general consensus that the PAEHR is more than a tool that serves for patient data collection and information exchange. It is a "third agent" during patient care encounters that essentially improves PCC [20,24]. For example, patients who used PAEHR portals prior to doctor visits reported that communication with their physicians improved considerably [25]. This is because the patient data in the PAEHR system enables providers to monitor patients' symptoms and medication adherence [26]. Physicians thus would spend much time and pay more attention to patients during clinical encounters [27]. Meanwhile, patients who used PAEHR portals perceived more PCC, as they felt empowered to ask questions or offer comments regarding their health problems [24,28]. By this token, PAEHR portal use and PCC can facilitate patients' management of their health and should eventually contribute to health improvement [20,21,29]. Street et al [29] proposed a pathway model of health communication and suggested that, in most cases, PCC affects patient health through a more indirect route via an intermediate outcome of communication, such as health self-efficacy. It is understandable that PCC can increase patients' health self-efficacy because providers' clear explanations and expressions of support could increase patient knowledge and shared understanding, motivate patients to follow through with treatment recommendations, and thus improve patients' confidence in self-care management.

Following this line, 2 mediators—PCC and health self-efficacy—were conceptualized as the proximal and intermediate outcomes of PAEHR portal use, respectively. Previous research that examined related variables has provided empirical support. For instance, Madhavan et al [30] found that due to the transportability and interoperability, effective use of PAEHR contributes to improved PCC, which plays a cardinal role in cancer survivors' health management. Guo et al [31] found that eHealth adoption (eg, seeking web-based health information and using health apps) was significantly associated with improved self-care skills, which further led to more positive self-rated health among Taiwanese patients with chronic diseases [31]. Liu and Yeo [22] conceptualized a framework, suggesting that web-based patient-provider communication via eHealth technologies may improve patients' quality of life through sequential mediators of patient-centered care and health management skills. Building on prior research, this study aims to examine the relationships among cancer survivors' PAEHR portal use, PCC, health self-efficacy, and health outcomes. Moreover, the mediation roles of PCC and health self-efficacy

were tested. Thus, the following direct and indirect relationships between PAEHR portal use and cancer survivors' health outcomes (see Figure 1) were proposed:

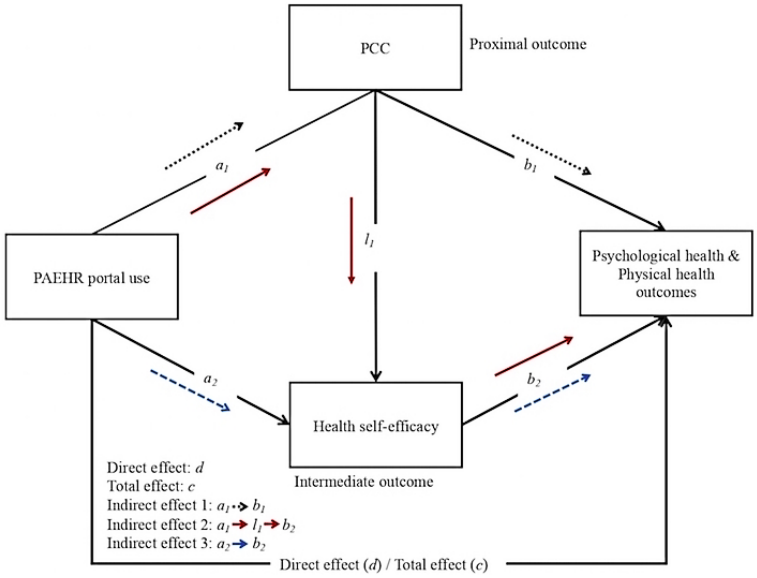
Hypothesis 1: PAEHR portal use is positively related to cancer survivors' health outcomes.

Hypothesis 2: PCC mediates the relationship between PAEHR portal use and cancer survivors' health outcomes.

Hypothesis 3: Health self-efficacy mediates the relationship between PAEHR portal use and cancer survivors' health outcomes.

Hypothesis 4: PCC and health self-efficacy sequentially mediate the relationship between PAEHR portal use and cancer survivors' health outcomes.

Figure 1. Pathways between patient-accessible electronic health record portal use and health outcomes. a_1, a_2, b_1, b_2, and l_1 indicate the pathways and the effects. PAEHR: patient-accessible electronic health record; PCC: patient-centered communication.



Methods

Study Design and Sample Population

Data for this study were derived from the Health Information National Trends Survey (HINTS 5, Cycle 4) collected from February 2020 to June 2020. HINTS is administered by the National Cancer Institute in the United States to collect nationally representative data about American adults' access to health-related information, health behaviors, and health outcomes. The survey design and sampling procedures for HINTS have been explicated extensively in previous research [32]. The final sample of HINTS 5, Cycle 4 consisted of 3865 respondents (response rate=36.7%) of the 10,531 participants. This study only involved respondents who reported having been diagnosed with cancer (N=626).

Ethical Considerations

This study used secondary data. The HINTS data meet strict ethical standards and have obtained ethics approval. Informed consent has been obtained from all participants, and all methods were carried out in accordance with relevant guidelines and regulations.

Measures

PAEHR portal use was measured by asking respondents whether they had accessed patient portals of PAEHR in the past year for certain eHealth activities [33]. Three items were included: "Look up test results," "securely message health care provider and staff," and "download health information to computer or mobile device." Responses were dichotomous (no=0, yes=1) and added up to represent PAEHR portal use (mean 1.726, SD 0.575).

PCC consisted of 7 statements that assessed patients' perceptions of communication with all doctors, nurses, or other health professionals in the past 12 months [21,34]. A 4-point Likert scale (1=always, 4=never) was used. Responses to the 7 statements were reversely coded and averaged to create the index of PCC, and higher values represent high levels of PCC (mean 3.414, SD 0.607; Cronbach α=.93).

Health self-efficacy was measured using 1 item to assess one's ability to take care of his/her health on a 5-point scale from 1 (completely confident) to 5 (not confident at all) [23]. Respondents' answers were reversely scored, and a higher score represented a higher level of health self-efficacy (mean 3.804, SD 0.812).

Physical health was measured by 4 items on comorbidities, drawn from prior research of similar measures [35]. Respondents were asked whether they had been told by a doctor or another health professional that they had medical conditions such as (1) diabetes or high blood sugar; (2) high blood pressure or hypertension; (3) a heart condition such as heart attack, angina, or congestive heart failure; and (4) chronic lung disease, asthma, emphysema, or chronic bronchitis. Responses to these items were dichotomous (no=0, yes=1). The answers were added up, and a higher value indicated better physical health (mean 2.748, SD 1.082).

Psychological health was measured by 4 items derived from previous research [36]. Sample items included "feeling down,

depressed, or hopeless" and "feeling nervous, anxious, or on edge." The 4 items were measured on a 4-point scale (1=nearly every day to 4=not at all) and averaged to form a composite score representing psychological health (mean 3.502, SD 0.706; Cronbach α=.88). A higher value suggests better psychological health. The descriptive details of the focal variables are shown in Tables 1-4.

The control variables included demographics such as age, gender (male=1, female=0), education (less than 8 years=1, postgraduate=7), annual household income (US $0-9999=1, US $200,000 or more=9), and race (non-Hispanic White=1, others=0).

Table 1. Descriptive statistics of the patient-accessible electronic health record portal use and physical health of the participants (N=626).

	Yes	No	Nonvalid
Patient-accessible electronic health record portal use, n (%)			
Look up test results	252 (40.3)	36 (5.8)	338 (53.9)
Securely message health care provider and staff	176 (28.1)	110 (17.6)	340 (54.3)
Download health information to computer or mobile device	68 (10.9)	218 (34.8)	340 (54.3)
Physical health, n (%)			
Diabetes or high blood sugar	176 (28.1)	440 (70.3)	10 (1.6)
High blood pressure or hypertension	374 (59.7)	244 (39)	8 (1.3)
A heart condition such as heart attack, angina, or congestive heart failure	91 (14.5)	527 (84.2)	8 (1.3)
Chronic lung disease, asthma, emphysema, or chronic bronchitis	132 (21.1)	486 (77.6)	8 (1.3)

Table 2. Descriptive statistics of patient-centered communication (N=626).

Patient-centered communication	Always, n (%)	Usually, n (%)	Sometimes, n (%)	Never, n (%)	Nonvalid, n (%)
Give you the chance to ask all the health-related questions you had	393 (62.8)	142 (22.7)	39 (6.2)	3 (0.5)	49 (7.8)
Give the attention you needed to your feelings and emotions	279 (44.6)	185 (29.6)	83 (13.3)	23 (3.7)	56 (8.9)
Involve you in decisions about your health care as much as you wanted	324 (51.8)	180 (28.8)	65 (10.4)	7 (1.1)	50 (7.9)
Make sure you understood the things you needed to do to take care of your health	362 (57.8)	169 (27)	43 (6.9)	3 (0.5)	49 (7.8)
Explain things in a way you could understand	366 (58.5)	164 (26.2)	43 (6.9)	3 (0.5)	50 (7.9)
Spend enough time with you	292 (46.6)	193 (30.8)	73 (11.7)	17 (2.7)	51 (8.2)
Help you deal with feelings of uncertainty about your health or health care	260 (41.5)	191 (30.5)	90 (14.4)	29 (4.6)	56 (9)

Table 3. Descriptive statistics of health self-efficacy (N=626).

Health self-efficacy	Completely confident	Very confident	Somewhat confident	A little confident	Not confident at all	Nonvalid
How confident are you about your ability to take good care of your health, n (%)	111 (17.7)	318 (50.8)	159 (25.4)	28 (4.5)	6 (1)	4 (0.6)

Table 4. Descriptive statistics of psychological health (N=626).

Psychological health	Nearly every day, n (%)	More than half the day, n (%)	Several days, n (%)	Not at all, n (%)	Nonvalid, n (%)
Little interest or pleasure in doing things	31 (5)	53 (8.5)	123 (19.6)	404 (64.5)	15 (2.4)
Feeling down, depressed, or hopeless	18 (2.9)	31 (5)	122 (19.5)	436 (69.6)	19 (3)
Feeling nervous, anxious, or on edge	30 (4.8)	29 (4.6)	163 (26)	389 (62.1)	15 (2.5)
Not being able to stop or control worrying	29 (4.6)	44 (7)	112 (17.9)	424 (67.7)	17 (2.8)

Data Analysis

Data analysis was performed using SPSS version 26 (IBM Corp). First, the MEAN () function was used to compute the mean of multiple-item variables that at least one item has a valid value or single-item variables that have valid values. Otherwise, the cases were considered missing in the following analysis. Besides, as a complementary technique, min-max normalization [37] was introduced to compare the estimates of all the paths in the mediation model. Specifically, all research variables were converted into a common measurement scale of 0 to 1. For example, we can subtract 1 from a 5-point rating to adjust the scale to start at 0 and then divide it by 4 to compress the scale. In this study, we referred to the regression coefficients generated by min-max normalization as percentage coefficients (b_p) [38,39]. Second, the mean substitution was used for all missing cases. Third, descriptive statistics was analyzed. Fourth, the mediation models were tested using Model 6 from the SPSS macro PROCESS; statistically significant relationships among PAEHR portal use, PCC, health self-efficacy, and physical and psychological health were examined using bootstrapping procedures. The 95% bootstrapped CIs were used with 10,000 resamplings.

Results

The mean age of the cancer survivors was 67.46 (SD 13.19; range 19-104) years. There were more female respondents (370/626, 59.1%) than male respondents (256/626, 40.9%). The majority of the participants had received some college education (405/626, 64.7%), were non-Hispanic White (428/626, 68.4%), and had annual household income between US $35,000 and US $74,999 (259/626, 41.4%). The detailed demographic information is summarized in Table 5.

Table 5. Sample population characteristics (N=626).

Characteristic	Value
Age in years, mean (SD)	67.46 (13.19)
Gender, n (%)	
Male	256 (40.9)
Female	370 (59.1)
Education, n (%)	
Less than 8 years of education	14 (2.2)
8-11 years of education	29 (4.6)
12 years of education or completed high school	132 (21.1)
Post high school training other than college	46 (7.3)
Some college	143 (22.8)
College graduate	145 (23.2)
Postgraduate	117 (18.7)
Annual income (USD), n (%)	
0-9999	33 (5.3)
10,000-14,999	34 (5.4)
15,000-19,999	37 (5.9)
20,000-34,999	79 (12.6)
35,000-49,999	87 (13.9)
50,000-74,999	172 (27.5)
75,000-99,999	58 (9.3)
100,000-199,999	94 (15)
200,000 or more	32 (5.1)
Race, n (%)	
Non-Hispanic White	428 (68.4)
Others	198 (31.6)

Hypothesis 1 posited that PAEHR portal use is positively related to cancer survivors' health outcomes. Table 6 shows that there was no significant direct association between PAEHR portal use and cancer survivors' health outcomes, irrespective of the physical or psychological health. Thus, hypothesis 1 was not supported.

Hypothesis 2 predicted that PCC mediates the relationship between PAEHR portal use and cancer survivors' health

outcomes. As depicted in Table 6, PAEHR portal use was significantly and positively associated with PCC (b_p=0.131; β=.125, 95% CI .048-.214; *P*=.002) in the 2 models. Meanwhile, PCC was positively associated with cancer survivors' psychological health (b_p=0.270; β=.269, 95% CI .258-.461; *P*<.001). No significant relationship between PCC and cancer survivors' physical health was acknowledged. The results indicated that PCC indeed mediated the relation between PAEHR portal use and cancer survivors' psychological health (b_p=0.029; β=.023, 95% CI .009-.054), whereas the counterpart effect failed to pass the statistical threshold (95% CI contained zero) for physical health. Hypothesis 2 was partially supported.

Hypothesis 3 predicted that PAEHR portal use might increase cancer survivors' health outcomes through the mediation of association with health self-efficacy. The mediation effects in the 2 models were statistically unacknowledged. Thus, hypothesis 3 was not supported.

Hypothesis 4 predicted that PAEHR portal use will be related to cancer survivors' health outcomes through the serial mediation of PCC and health self-efficacy. As shown in Table 6, the indirect relationship between PAEHR portal use and cancer survivors' physical health (b_p=0.006; β=.004, 95% CI .002-.018) and between PAEHR portal use and psychological health (b_p=0.006; β=.005, 95% CI .002-.014) via sequential mediators of PCC and health self-efficacy were statistically acknowledged, thereby supporting hypothesis 4.

Table 6. Mediation models[a].

	b_p[b]	β	SE	95% CI	*P* value[c]
Dependent variable: Psychological health (Model 1)					
PAEHR[d]→PCC[e] (a_1 path)	0.131	.125	.042	.048 to .214	.002
PAEHR→Health self-efficacy (a_2 path)	0.022	.021	.055	–.078 to .137	.59
PCC→Health self-efficacy (l_1 path)	0.270	.269	.052	.258 to .461	<.001
PCC→Psychological health (b_1 path)	0.217	.186	.046	.127 to .306	<.001
Health self-efficacy→Psychological health (b_2 path)	0.181	.156	.034	.068 to .202	<.001
PAEHR→Psychological health (direct effect, *d* path)	–0.016	–.013	.046	–.108 to .075	.73
PAEHR→Psychological health (total effect, *c* path)	0.023	.018	.048	–.072 to .117	.64
PAEHR→PCC→ Psychological health (indirect effect, a_1xb_1)	0.029	.023	.012	.009 to .054	N/A[f]
PAEHR→PCC→ Health self-efficacy→Psychological health (indirect effect, a_1xb_2xl_1)	0.006	.005	.003	.002 to .014	N/A
PAEHR→Health self-efficacy→Psychological health (indirect effect, a_2xb_2)	0.004	<.001	.008	–.012 to .020	N/A
Dependent variable: Physical health (Model 2)					
PAEHR→PCC (a_1 path)	0.131	.125	.042	.048 to .214	.002
PAEHR→Health self-efficacy (a_2 path)	0.022	.021	.055	–.078 to .137	.59
PCC→Health self-efficacy (l_1 path)	0.270	.269	.052	.258 to .461	<.001
PCC→Physical health (b_1 path)	0.013	.010	.070	–.120 to .154	.81
Health self-efficacy→Physical health (b_2 path)	0.168	.126	.052	.066 to .270	.001
PAEHR→Physical health (direct effect, *d* path)	–0.032	–.023	.071	–.183 to .096	.55
PAEHR→Physical health (total effect, *c* path)	–0.021	–.015	.071	–.168 to .112	.69
PAEHR→PCC→Physical health (indirect effect, a_1→b_1)	0.002	.001	.011	–.020 to .024	N/A
PAEHR→PCC→Health self-efficacy→Physical health (indirect effect, a_1→b_2→l_1)	0.006	.004	.004	.002 to .018	NA
PAEHR→Health self-efficacy→Physical health (indirect effect, a_2→b_2)	0.004	.003	.010	–.015 to .026	N/A

[a]a_1, a_2, b_1, b_2, and l_1 in this table indicate the pathways between patient-accessible electronic health record portal use and health outcomes and the effects.

[b]Regression coefficient generated by min-max normalization as percentage coefficient.

[c]*P* values are not computed for bootstrapped indirect effects.

[d]PAEHR: patient-accessible electronic health record.

[e]PCC: patient-centered communication.

[f]N/A: not applicable.

Discussion

Principal Findings

In light of the existing literature on the robust salutary effects of PAEHR portals on patient health, our study examined the effects of PAEHR portal use on cancer survivors' health outcomes as well as the mediating roles of PCC and health self-efficacy. The results of our study indicated that the significant effect of PAEHR portal use on cancer survivors' physical and psychological health was indirect through the mediated associations with PCC and health self-efficacy.

The direct association between PAEHR portal use and cancer survivors' health outcomes is not acknowledged in this study. The findings of our study emphasize the mediation mechanisms through which the PAEHR portal use exerts an influence on cancer survivors' physical and psychological health, which were in accordance with that reported in previous research that theorizes the process through which PAEHR may impact patient health [20]. Rathert et al's [20] and Street et al's [29] pathway models provide the needed theoretical foundation for this study, supporting that several steps must occur for health improvement to be influenced by cancer survivors' PAEHR portal use. First, PAEHR portals serve as a tool that facilitates patient-provider communication. Physicians should incorporate PAEHR systems to provide PCC that supports patients in making informed health care decisions that are consistent with their needs, values, and preferences. Unless PCC is improved, PAEHR portal use will not increase patients' health self-efficacy and improve their health outcomes. Although previous research has identified the association between PAEHR and patient health, we investigated the mediating mechanisms (the process) through which PAEHR impacts patient health.

PCC and health self-efficacy were identified as the intrinsic and extrinsic factors of PAEHR, respectively, that help explain how PAEHR portal use influences patients' health outcomes. The results of our study suggest that PCC can partially mediate the relationship between PAEHR portal use and cancer survivors' psychological health. The mediation results indicated that the more cancer survivors use the PAEHR portals to stay informed about their health and communicate with health care professionals, the more likely they are to perceive PCC, which in turn results in more positive psychological health. A plausible reason is that the increasing accessibility to health professionals and patient information facilitated by PAEHR systems may enhance patient involvement in their health care decision-making [40]. Through PAEHR portals, cancer survivors are likely to be informed about their health status, be well educated with adequate health information, and have convenient access to health care professionals for medical guidance [41]. As a result, patients feel more engaged in PCC, which helps better understand their health and motivate them to stay positive and improve their psychological health [42-44]. However, PCC has no mediation effect between PAEHR portal use and cancer survivors' physical health. This might be because the research sample of this study consisted of 626 cancer survivors with an average age >60 years, and they were likely to have inferior health status. PCC could not improve physical health unless patients were equipped with the necessary health skills. This assumption was supported by the sequential mediation effect of PCC and health self-efficacy between PAEHR portal use and cancer survivors' health outcomes.

The results of our study showed that PCC is positively associated with health self-efficacy, and higher levels of health self-efficacy can enhance cancer survivors' physical and psychological health. This finding was consistent with prior research, suggesting that PCC may empower patients, help increase their self-care skills, and provide the needed information and support to facilitate patients' health management [45,46]. Furthermore, improved health self-efficacy can help people take care of their physical and psychological health, and this finding was congruent with previous findings [47-49]. Our results provide empirical evidence of the indirect effect of PAEHR portal use on cancer survivors' health outcomes through PCC and health self-efficacy.

Comparison With Prior Work

Our study in comparison with previous work has heuristic value for public health research in several ways. First, the findings of our study offer empirical support for eCCM [19] and Rathert et al's [20] pathway model in understanding the process through which PAEHR impacts patient health. Second, this study extends the current literature by investigating the usability of eHealth technologies in delivering longitudinal survivorship care for patients with chronic diseases as well as examining the mediation roles of PCC and health self-efficacy. Our findings stressed PCC as the salient intrinsic factor of PAEHR that helps improve patients' health self-efficacy and prompts them into action to maintain their health. The mediation effects provide a more nuanced understanding of the mechanisms underlying the association between PAEHR portal use and patients' health outcomes. This model was established in several hypotheses by which the assumptions have been shown tenable. This study thus helps consolidate past research on the relationships between PAEHR portal use and patients' physical and psychological health.

This study also has important practical implications. First, given the important role of electronic means for health management, multifaceted strategies should be implemented to promote the assimilation of PAEHR at both institutional and individual levels. For example, through patient education and support, patients can gain knowledge about PAEHR and be encouraged to integrate PAEHR into their health care in everyday life. Besides, we should also encourage medical professionals to engage in PAEHR systems to provide customized health care services. For example, a medical professional can provide detailed explanations for certain clinical decisions through PAEHR portals, and patients can access and revisit the messages that can facilitate their self-care practices [50]. Second, considering the significant role of PAEHR portals, we should continue to develop information technology infrastructure to improve the accessibility of high-quality and long-term survivorship care. For example, patients who live remotely with low-speed internet and people who have poor internet skills may not benefit from the convenience and great efficiency brought by the internet for medical consultations [47]. Thus,

information and communication technology companies should expand high-speed internet provisions to the other regions and deliver benefits to more people and communities. In addition, we should provide continuous support to help individuals overcome the barriers encountered in using PAEHR portals for health management [51]. Third, strict policies for web-based health service regulation should be implemented to protect patients' information and to ensure a safe PAEHR environment. In parallel with the governmental measures, it is equally important to educate patients about their rights to access health data and responsibilities for personal information security. Fourth, considering the effect of PCC, it is important to help patients more actively participate in health consultations as well as provide training to physicians in delivering empathetic, mindful, informative, and patient-centered care.

Limitations and Directions for Future Research

Several limitations of this study should be noted. First, owing to the cross-sectional design of HINTS, we know little about the causal inferences of relationships examined in this study. Further research should collect panel data or use experimental research designs to better understand the relationships among PAEHR portal use, PCC, health self-efficacy, and health outcomes. Second, according to CCM and eCCM, there are 6 key components of eHealth technologies for care delivery, such as health system support and delivery system design. However, PAEHR portal use in this study was measured using 3 items, that is, patients' past experience in PAEHR portal use for checking test results, patient-provider communication, and health information acquisition. We know little about the influence of other aspects of PAEHR portal use. To our knowledge, no study has examined the usability of PAEHR system design and how it impacts patient-provider communication and patients' health maintenance. Besides, PAEHR portal use was examined as an integrated concept, and we hardly know how different types of PAEHR portal usage may affect patient health differently. Based on this study, future research should take into account the different use dimensions of PAEHR systems or the different types of PAEHR portal usage and compare their different influences. Third, PCC and health self-efficacy were identified as the mediators in the relationship between PAEHR portal use and cancer survivors' health outcomes. Other potential interveners might be overlooked. Researchers should further extend the model and identify other mediators (eg, knowledge) or moderators (eg, health literacy, digital literacy) that significantly influence PAEHR portal users' health-related outcomes. Fourth, the research findings of our study might be impacted by sampling bias. For example, more than half of the respondents were aged between 60 years and 80 years (mean 67.46 years) and had at least completed some college education. It is recommended that a more representative sample be analyzed to better understand the full range of cancer survivors' PAEHR portal use. Moreover, our study focused on cancer survivors, and the results may not be generalizable to other populations. PAEHR portals can likely be helpful and useful for people with other chronic conditions such as diabetes and asthma. Thus, researchers should replicate this work in other populations to obtain more tentative evidence, thereby supporting the positive association between PAEHR portal use and health outcomes.

Conclusion

This study offers empirical evidence on the influence of PAEHR portal use on cancer survivors' physical and psychological health. Although electronic technologies have been widely applied in health care settings, the adoption rate of PAEHR among patients remains low. This study suggests that PAEHR portal use is vital in delivering longitudinal survivorship care for cancer survivors. In particular, the influence of PAEHR portal use on health outcomes may be indirect through the mediated associations with PCC care and health self-efficacy. Understanding these relationships can help increase the use of PAEHR portals, promote PCC, enhance patients' health self-efficacy, and eventually improve their physical and psychological health.

Acknowledgments

This research was supported in part by grants of the University of Macau, ICI-RTO-0010-2021, CPG20XX-00035-FSS, and SRG20XX-00143-FSS, and a grant of Macau Higher Education Fund (HSS-UMAC-2020-02).

Data Availability

The data sets generated and analyzed during this study are publicly available on [52]. All data and materials comply with field standards.

Conflicts of Interest

None declared.

References

1. Cancer. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/cancer [accessed 2021-06-30]
2. Cancer facts and figures. American Cancer Society. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf [accessed 2021-07-08]
3. Naughton M, Weaver K. Physical and mental health among cancer survivors: considerations for long-term care and quality of life. N C Med J 2014;75(4):283-286 [FREE Full text] [doi: 10.18043/ncm.75.4.283] [Medline: 25046097]

4. Denlinger C, Ligibel J, Are M, Baker KS, Demark-Wahnefried W, Friedman DL, National Comprehensive Cancer Network. Survivorship: sleep disorders, version 1.2014. J Natl Compr Canc Netw 2014 May;12(5):630-642 [FREE Full text] [doi: 10.6004/jnccn.2014.0067] [Medline: 24812132]
5. Keating N, Nørredam M, Landrum M, Huskamp H, Meara E. Physical and mental health status of older long-term cancer survivors. J Am Geriatr Soc 2005 Dec;53(12):2145-2152. [doi: 10.1111/j.1532-5415.2005.00507.x] [Medline: 16398900]
6. Epstein RM, Street RL. Patient-centered communication in cancer care: promoting healing and reducing suffering. National Cancer Institute. URL: https://cancercontrol.cancer.gov/brp/docs/pcc_monograph.pdf [accessed 2021-07-08]
7. Gorin S, Haggstrom D, Han P, Fairfield K, Krebs P, Clauser S. Cancer Care Coordination: a Systematic Review and Meta-Analysis of Over 30 Years of Empirical Studies. Ann Behav Med 2017 Aug;51(4):532-546. [doi: 10.1007/s12160-017-9876-2] [Medline: 28685390]
8. Singh H, Fox S, Petersen N, Shethia A, Street R. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009 Jun 16;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]
9. Patel M, Smith A, Leo H, Hao W, Zheng K. Improving Patient-Provider Communication and Therapeutic Practice Through Better Integration of Electronic Health Records in the Exam Room: A Pilot Study. Health Educ Behav 2019 Jun;46(3):484-493 [FREE Full text] [doi: 10.1177/1090198118796879] [Medline: 30196720]
10. Abramson E, Barrón Y, Quaresimo J, Kaushal R. Electronic Prescribing Within an Electronic Health Record Reduces Ambulatory Prescribing Errors. The Joint Commission Journal on Quality and Patient Safety 2011 Oct;37(10):470-478 [FREE Full text] [doi: 10.1016/s1553-7250(11)37060-2]
11. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008 Oct 06;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
12. Fernández-Alemán JL, Señor IC, Lozoya P, Toval A. Security and privacy in electronic health records: a systematic literature review. J Biomed Inform 2013 Jun;46(3):541-562 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.003] [Medline: 23305810]
13. Kruse C, Stein A, Thomas H, Kaur H. The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. J Med Syst 2018 Sep 29;42(11):214 [FREE Full text] [doi: 10.1007/s10916-018-1075-6] [Medline: 30269237]
14. Wani D, Malhotra M. Does the meaningful use of electronic health records improve patient outcomes? Journal of Operations Management 2018 Jun 22;60(1):1-18 [FREE Full text] [doi: 10.1016/j.jom.2018.06.003]
15. Hossain A, Quaresma R, Rahman H. Investigating factors influencing the physicians' adoption of electronic health record (EHR) in healthcare system of Bangladesh: An empirical study. International Journal of Information Management 2019 Feb;44:76-87 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2018.09.016]
16. Barrett A. Electronic Health Record (EHR) Organizational Change: Explaining Resistance Through Profession, Organizational Experience, and EHR Communication Quality. Health Commun 2018 Apr;33(4):496-506. [doi: 10.1080/10410236.2016.1278506] [Medline: 28157382]
17. 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]
18. Siminerio L. The role of technology and the chronic care model. J Diabetes Sci Technol 2010 Mar 01;4(2):470-475 [FREE Full text] [doi: 10.1177/193229681000400229] [Medline: 20307409]
19. Gee P, Greenwood D, Paterniti D, Ward D, Miller L. The eHealth Enhanced Chronic Care Model: a theory derivation approach. J Med Internet Res 2015 Apr 01;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
20. Rathert C, Mittler J, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: What does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]
21. Liu P, Jiang S. Patient-Centered Communication Mediates the Relationship between Health Information Acquisition and Patient Trust in Physicians: A Five-Year Comparison in China. Health Commun 2021 Feb;36(2):207-216. [doi: 10.1080/10410236.2019.1673948] [Medline: 31617412]
22. Liu P, Yeo T. How Online Patient-Provider Communication Impacts Quality of Life: Examining the Role of Patient-Centered Care and Health Competence. Health Commun 2021 Aug 02:1-6. [doi: 10.1080/10410236.2021.1961971] [Medline: 34340609]
23. Zou W. How Do Patient-Centered Communication, Emotional Well-Being, Confidence in Health Information-Seeking, and Future Time Perspective Influence Health Self-Efficacy? American Journal of Health Education 2021 Apr 08;52(3):137-144 [FREE Full text] [doi: 10.1080/19325037.2021.1902887]
24. Pearce C, Kumarapeli P, de Lusignan S. "Effects of exam room EHR use on doctor-patient communication: a systematic literature review" - triadic and other key terms may have identified additional literature. Inform Prim Care 2013;21(1):40-42 [FREE Full text] [doi: 10.14236/jhi.v21i1.39] [Medline: 24629655]
25. Wald J, Businger A, Gandhi T, Grant R, Poon E, Schnipper J, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]
26. Young H, Nesbitt T. Increasing the Capacity of Primary Care Through Enabling Technology. J Gen Intern Med 2017 Apr;32(4):398-403 [FREE Full text] [doi: 10.1007/s11606-016-3952-3] [Medline: 28243871]

27. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
28. McGrath J, Arar N, Pugh J. The influence of electronic medical record usage on nonverbal communication in the medical interview. Health Informatics J 2007 Jun;13(2):105-118 [FREE Full text] [doi: 10.1177/1460458207076466] [Medline: 17510223]
29. Street RJ, Makoul G, Arora N, Epstein R. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
30. Madhavan S, Sanders AE, Chou WS, Shuster A, Boone KW, Dente MA, et al. Pediatric palliative care and eHealth opportunities for patient-centered care. Am J Prev Med 2011 May;40(5 Suppl 2):S208-S216 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.013] [Medline: 21521596]
31. Guo S, Hsing H, Lin J, Lee C. Relationships Between Mobile eHealth Literacy, Diabetes Self-care, and Glycemic Outcomes in Taiwanese Patients With Type 2 Diabetes: Cross-sectional Study. JMIR Mhealth Uhealth 2021 Feb 05;9(2):e18404 [FREE Full text] [doi: 10.2196/18404] [Medline: 33544088]
32. Park C, Cho D, Moore P. How does education lead to healthier behaviours? Testing the mediational roles of perceived control, health literacy and social support. Psychol Health 2018 Nov;33(11):1416-1429. [doi: 10.1080/08870446.2018.1510932] [Medline: 30450977]
33. Hong Y, Jiang S, Liu P. Use of Patient Portals of Electronic Health Records Remains Low From 2014 to 2018: Results From a National Survey and Policy Implications. Am J Health Promot 2020 Jul;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
34. Blanch-Hartigan D, Chawla N, Moser R, Finney Rutten LJ, Hesse B, Arora N. Trends in cancer survivors' experience of patient-centered communication: results from the Health Information National Trends Survey (HINTS). J Cancer Surviv 2016 Dec;10(6):1067-1077. [doi: 10.1007/s11764-016-0550-7] [Medline: 27193357]
35. Greenberg A, Falisi A, Finney Rutten LJ, Chou WYS, Patel V, Moser RP, et al. Access to Electronic Personal Health Records Among Patients With Multiple Chronic Conditions: A Secondary Data Analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
36. Ellis E, Orom H, Giovino G, Kiviniemi M. Relations between negative affect and health behaviors by race/ethnicity: Differential effects for symptoms of depression and anxiety. Health Psychol 2015 Sep;34(9):966-969. [doi: 10.1037/hea0000197] [Medline: 25581702]
37. Jayalakshmi T, Santhakumaran A. Statistical Normalization and Back Propagationfor Classification. IJCTE 2011:89-93 [FREE Full text] [doi: 10.7763/ijcte.2011.v3.288]
38. Zhao X, Zhang X. Emerging methodological issues in quantitative communication research. New Trends Commun Stud. II 2014:953-978 [FREE Full text]
39. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis, First Edition: A Regression-Based Approach. eBook: The Guilford Press; 2022.
40. Roter D, Larson S, Sands D, Ford D, Houston T. Can e-mail messages between patients and physicians be patient-centered? Health Commun 2008;23(1):80-86. [doi: 10.1080/10410230701807295] [Medline: 18443995]
41. Zanaboni P, Kummervold P, Sørensen T, Johansen M. Patient Use and Experience With Online Access to Electronic Health Records in Norway: Results From an Online Survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
42. Calfas K, Sallis J, Zabinski M, Wilfley DE, Rupp J, Prochaska JJ, et al. Preliminary evaluation of a multicomponent program for nutrition and physical activity change in primary care: PACE+ for adults. Prev Med 2002 Feb;34(2):153-161. [doi: 10.1006/pmed.2001.0964] [Medline: 11817910]
43. Lee Y, Lin J. Do patient autonomy preferences matter? Linking patient-centered care to patient-physician relationships and health outcomes. Soc Sci Med 2010 Nov;71(10):1811-1818. [doi: 10.1016/j.socscimed.2010.08.008] [Medline: 20933316]
44. Rathert C, Wyrwich M, Boren S. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
45. Finney Rutten LJ, Hesse B, St Sauver JL, Wilson P, Chawla N, Hartigan DB, et al. Health Self-Efficacy Among Populations with Multiple Chronic Conditions: the Value of Patient-Centered Communication. Adv Ther 2016 Aug;33(8):1440-1451 [FREE Full text] [doi: 10.1007/s12325-016-0369-7] [Medline: 27357639]
46. Saha S, Beach M. The impact of patient-centered communication on patients' decision making and evaluations of physicians: a randomized study using video vignettes. Patient Educ Couns 2011 Sep;84(3):386-392 [FREE Full text] [doi: 10.1016/j.pec.2011.04.023] [Medline: 21600723]
47. Bachmann J, Goggins K, Nwosu S, Schildcrout J, Kripalani S, Wallston K. Perceived health competence predicts health behavior and health-related quality of life in patients with cardiovascular disease. Patient Educ Couns 2016 Dec;99(12):2071-2079 [FREE Full text] [doi: 10.1016/j.pec.2016.07.020] [Medline: 27450479]
48. Silva WD, Campos J, Marôco J. Impact of inherent aspects of body image, eating behavior and perceived health competence on quality of life of university students. PLoS One 2018;13(6):e0199480 [FREE Full text] [doi: 10.1371/journal.pone.0199480] [Medline: 29933390]

49. Kujala S, Hörhammer I, Väyrynen A, Holmroos M, Nättiaho-Rönnholm M, Hägglund M, et al. Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-sectional Survey. J Med Internet Res 2022 Jun 06;24(6):e37438 [FREE Full text] [doi: 10.2196/37438] [Medline: 35666563]
50. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo P. Electronic health literacy and health-related outcomes among older adults: A systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]
51. Vollbrecht H, Arora V, Otero S, Carey K, Meltzer D, Press V. Evaluating the Need to Address Digital Literacy Among Hospitalized Patients: Cross-Sectional Observational Study. J Med Internet Res 2020 Jun 04;22(6):e17519 [FREE Full text] [doi: 10.2196/17519] [Medline: 32496196]
52. Health information national trends survey (HINTS). National Cancer Institute. URL: https://hints.cancer.gov [accessed 2022-10-19]

Abbreviations

CCM: Chronic Care Model
eCCM: eHealth enhanced Chronic Care Model
HINTS: Health Information National Trends Survey
PAEHR: patient-accessible electronic health record
PCC: patient-centered communication

Edited by T Leung; submitted 17.05.22; peer-reviewed by R Yang, J Hagström; comments to author 13.08.22; revised version received 04.09.22; accepted 24.09.22; published 24.10.22.

Please cite as:
Liu PL, Zhao X, Ye JF
The Effects of the Use of Patient-Accessible Electronic Health Record Portals on Cancer Survivors' Health Outcomes: Cross-sectional Survey Study
J Med Internet Res 2022;24(10):e39614
URL: https://www.jmir.org/2022/10/e39614
doi:10.2196/39614
PMID:36279157

Original Paper

Interactivity, Quality, and Content of Websites Promoting Health Behaviors During Infancy: 6-Year Update of the Systematic Assessment

Danielle Jawad[1,2,3], BSc, MPH; Heilok Cheng[3,4,5], BSc, MND; Li Ming Wen[1,2,3,5], MD, MMED, PhD; Chris Rissel[1,6], BSc, MPH, PhD; Louise Baur[1,3,7], MBBS, PhD; Seema Mihrshahi[8], BSc, MPH, PhD; Sarah Taki[1,2,3,5], BSc, MND, PhD

[1]Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Health Promotion Unit, Population Health Research & Evaluation Hub, Sydney Local Health District, Sydney, Australia

[3]National Health and Medical Research Council Centre of Research Excellence in the Early Prevention of Obesity in Childhood - Translate, The University of Sydney, Sydney, Australia

[4]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[5]Sydney Institute for Women, Children and their Families, Sydney Local Health District, Sydney, Australia

[6]College of Medicine and Public Health, Rural and Remote Health South Australia and Northern Territory, Flinders University, Darwin, Australia

[7]Specialty of Child and Adolescent Health, Sydney Medical School, The University of Sydney, Sydney, Australia

[8]Department of Health Sciences, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, Australia

Corresponding Author:
Danielle Jawad, BSc, MPH
Sydney School of Public Health
Faculty of Medicine and Health
The University of Sydney
A27 Fisher road, Camperdown
Sydney, 2006
Australia
Phone: 61 2 9515 9895
Email: danielle.jawad@sydney.edu.au

Abstract

Background: As of 2021, 89% of the Australian population are active internet users. Although the internet is widely used, there are concerns about the quality, accuracy, and credibility of health-related websites. A 2015 systematic assessment of infant feeding websites and apps available in Australia found that 61% of websites were of poor quality and readability, with minimal coverage of infant feeding topics and lack of author credibility.

Objective: We aimed to systematically assess the quality, interactivity, readability, and comprehensibility of information targeting infant health behaviors on websites globally and provide an update of the 2015 systematic assessment.

Methods: Keywords related to infant milk feeding behaviors, solid feeding behaviors, active play, screen time, and sleep were used to identify websites targeting infant health behaviors on the Google search engine on Safari. The websites were assessed by a subset of the authors using predetermined criteria between July 2021 and February 2022 and assessed for information content based on the Australian Infant Feeding Guidelines and National Physical Activity Recommendations. The Suitability Assessment of Materials, Quality Component Scoring System, the Health-Related Website Evaluation Form, and the adherence to the Health on the Net code were used to evaluate the suitability and quality of information. Readability was assessed using 3 web-based readability tools.

Results: Of the 450 websites screened, 66 were included based on the selection criteria and evaluated. Overall, the quality of websites was mostly adequate. Media-related sources, nongovernmental organizations, hospitals, and privately owned websites had the highest median quality scores, whereas university websites received the lowest median score (35%). The information covered within the websites was predominantly poor: 91% (60/66) of the websites received an overall score of ≤74% (mean 53%, SD 18%). The suitability of health information was mostly rated adequate for literacy demand, layout, and learning and motivation of readers. The median readability score for the websites was grade 8.5, which is higher than the government recommendations (<grade 8). Overall, 74% (49/66) of the websites obtained a poor rating for interactivity, measuring active control, 2-way

communication, and synchronicity. The most common features found on websites were social media links (61/66, 92%), frequently asked questions (48/66, 73%), and videos (44/66, 67%). Only 14% (9/66) of websites presented culturally responsive information.

Conclusions: Quality, content, readability, and interactivity of websites promoting health behaviors during infancy ranged between poor and adequate. Since the 2015 systematic assessment, there was a slight improvement in the quality of websites but no difference in the Suitability Assessment of Materials rating and readability of information. There is a need for researchers and health care providers to leverage innovative web-based platforms to provide culturally competent evidence-based information based on government guidelines that are accessible to those with limited English proficiency.

(J Med Internet Res 2022;24(10):e38641) doi:10.2196/38641

KEYWORDS

breastfeeding; bottle feeding; websites; web-based platform; infant food; readability; accuracy; consumer; health information; interactivity; solid food; quality; grading; comprehensibility; infant; baby; babies; feeding; food; eating; nutrition; health behavior; web-based information; health website; sleep; screen time; rating

Introduction

Background

With technological advances and developments, internet access continues to increase [1]. Globally, approximately 4.53 billion people have access to web portals [2,3], with more than half using mobile devices and 38.5% using desktop computers to access the internet worldwide [4]. Increasingly, internet users are making use of the availability of web-based resources, with approximately 4.5% of all internet searches looking for health-related information [5-7]. Recently, nationwide lockdowns, social distancing, and restrictions due to the COVID-19 pandemic have led to an inevitable surge in internet use among individuals, including parents of young children, to seek health information on the web [8,9] and by health care practitioners to assist in service delivery [10,11]. Given that the internet offers considerable opportunities for immediate and easy access to web-based resources, it has become a significant medium for the dissemination of health-related information.

A universal growing demand for, and use of, web-based resources related to child health information is evident [12-15]. The consolidated behavior of using web-based resources for health information is particularly common among new and expecting parents, where they most frequently search the internet for information related to infant nutrition, development, social support, and health symptoms [16-22]. A 2016 Australian survey found that more than 73% of parents with children aged <5 years used websites and web-based forums to access child health–related information [23]. Interestingly, 30% of those parents reported not trusting the information sources [23]. Another recent study in Switzerland showed that 91% of parents with at least one child aged <2 years used the internet to search for information related to their child's health and development [24]. The most frequently used sources reported were search engines (55%) and websites for parents (47%) [24]. Although the majority used the internet to search for health-related information, a large percentage of parents were skeptical about the trustworthiness of the web-based resources and their ability to correctly interpret the reliability of the health information they found [24]. This highlights the need for, and importance of, accessible websites to present health information accurately while ensuring it can be easily understood by their intended users.

This study updates and expands on a 2015 systematic assessment of infant feeding websites and mobile apps available in Australia [25]. The update of apps has been recently conducted by Cheng et al [26] in 2020. Hence, this study focused on updating and expanding the assessment of websites globally. The 2015 assessment found that 61% of Australian websites were of poor quality, with minimal coverage of infant feeding topics, lack of author credibility, and abstruse readability of content [25]. Since the publication of the 2015 systematic assessment, several other website assessments reviewing information related to infant health behaviors have identified similar findings [27,28]. A Korean study reported that websites were scored poorly when evaluated for availability, quality, and reliability of infant health information on the web [27]. Moreover, a recent analysis of 197 websites addressing preterm infants' health information also found that the overall quality of websites was low to moderate in terms of reliability and content [28]. Provision of inadequate or incomplete infant health information on the web could result in parental confusion, apprehension, and poorer care for infants when parents are unable to evaluate the accuracy and credibility of the web-based information.

Over the past 10 years, websites have evolved from static *read only* to a more interactive and fully immersive experience [29,30]. Since the 2015 assessment of websites [25], there has been a marked growth in bandwidth levels enabling the design of more sophisticated websites to offer consumer-oriented health information in various interactive ways, such as videos, parent forums, podcasts, and multilingual options that have provided a context to support culturally diverse people across the world [31-38]. In addition, the emergence of artificial intelligence and machine learning since 2016 has given rise to chatbot technology that simulates human-like conversations to provide consumers with support and relevant information [39]. This has been tested and proven to be successful among parents of young infants searching for information related to infant sleep and feeding practices [40].

Objective

With more parents resorting to web-based sources to seek infant health information, it is imperative for resources on the internet to reflect the latest infant and child health guidelines. Therefore, the aim of this study was to update the 2015 systematic assessment of websites [25] by evaluating the content, suitability, readability, comprehensibility, and quality of

information targeting infant nutrition, active play, screen time, and sleep behaviors on websites globally. In addition, this review expanded the 2015 systematic assessment [25] by examining interactivity, features, and cultural considerations of the websites.

Methods

Study Design

A systematic search and assessment were conducted to identify and evaluate websites targeting infant feeding, active play, screen time, and sleep behaviors between July 2021 and February 2022. As shown in Table 1, a range of validated tools was used to assess the selected websites. Details of the evaluation tools are described in Multimedia Appendix 1, and the details of the methods are given in Multimedia Appendix 2.

Table 1. Comparison of the systematic assessment between the 2015 assessment and this study.

Criteria	Systematic assessment	
	2015	2021
Website selection		
Australian websites only	✓	
Global websites including Australian websites		✓
Topic areas		
Milk feeding practices (breastfeeding and formula)	✓	✓
Solid feeding behaviors	✓	✓
Infant active play		✓
Infant screen time		✓
Infant sleep		✓
Scope, accuracy, and depth of information		
Excel spreadsheet built with an assessment criterion of 8 topics and 22 subtopics	✓	
Comprehensive REDCap[a] tool built with an assessment criterion of 9 topics and 65 subtopics		✓
Quality assessment		
Quality Component Scoring System	✓	✓
Health-Related Website Evaluation Form	✓	✓
Adherence to the Health on the Net code	✓	✓
Suitability of information		
The Suitability Assessment of Material	✓	✓
Readability		
Flesch-Kincaid	✓	✓
Simple Measure of Gobbledygook	✓	✓
Consensus based on 7 readability formulas		✓
Website interactivity and features		
The interactivity scale (15 items)		✓
Interactive features on websites		✓
Addresses culture		✓

[a]REDCap: Research Electronic Data Capture.

Stage 1: Website Selection

Overview

Websites were identified using the Chrome browser and Google search engine. All cookies and search history were erased from the web browser to ensure no previous web-based activities influenced the search results. The search terms were *Infant feeding, Baby food, Breast feeding, Infant feeding to appetite, Infant formula feeding, Introducing solid foods to baby, Good foods to start baby with no teeth, Best puree for babies, Solids and fussy babies, Solids and milk feeding, Infant active play, Tummy time, Screen time, Infant sleep, Baby co-sleep.* These

key terms were identified from questions asked in Facebook groups that consisted of parents with infants and from "related searches" on Google, which was used as a cross-reference to ensure the representativeness of the keywords.

Evidence shows that users concentrate their exploration of websites on the first 10 search results retrieved from a search engine and rarely go beyond the first 2 pages [41]. Hence, the first 30 websites generated from every search term were screened.

Inclusion Criteria

We included global websites that used English as a primary language or language option, were free of charge, targeted at parents of infants, and contained information on at least one of the following topics: milk feeding behaviors (breastfeeding, formula feeding, expressing breast milk, feeding to appetite, frequency or timing of feeding, and correct preparation of infant formula, storage of milk, quantity of milk, and transport of milk), solid food feeding behaviors (age of introduction, types of food introduced, and food allergies), infant activity ("tummy time," infant play, and movement), and infant screen time or infant sleep (bedtime routine, recommended hours of sleep, and cosleep) regardless of whether the websites addressed other content or age groups.

Exclusion Criteria

We excluded websites that had no information on one of the topics of interest listed in the inclusion criteria; were newspapers; were electronic books; required a password; had a payment fee; or had a link that redirected user to a scientific article, podcast, or downloadable Microsoft Word document and PDF document.

The first author screened all the websites for eligibility using predefined inclusion and exclusion criteria. Any uncertainties or disagreements regarding the inclusion of websites in the study were cross-checked by researchers ST, LMW, LB, and CR in a group meeting and discussed until consensus was reached.

Stage 2: Website Evaluation

Scope, Accuracy, and Depth of Information

Scope, accuracy, and depth of information were evaluated using a newly built tool on a password-protected database (REDCap [Research Electronic Data Capture]; Vanderbilt University) that was based on the Australian government's guidelines on infant feeding [42], physical activity [43], and sleep [44]. The tool consists of 11 broad topics with 65 subtopics on encouraging and supporting breastfeeding; initiating, establishing, and maintaining breastfeeding; management of common breastfeeding problems; expressing and storing breast milk; breastfeeding in specific situations; infant formula; solid food introduction; encouraging infant active play; screen time; and infant sleep behaviors. Each subtopic was scored as correct (+1), incorrect (−1), not addressed (0), or not applicable (which was not counted in the denominator of the overall score). For subtopics that were partially addressed, a partially complete (+0.5) score was given. A summary section score was automatically calculated for each topic, and a final overall score was generated after the assessment of all the content for the 11 topic areas. Overall scores were summarized as excellent (≥90%), adequate (75%-89%), or poor (≤74%) using the criteria from the Health-Related Website Evaluation Form (HRWEF) [45] similar to the updated assessment of apps [26].

Website Quality

Website quality was evaluated using the same validated tools as the 2015 assessment: the Quality Component Scoring System (QCSS) [46,47], the HRWEF [45], and the adherence to the Health on the Net Foundation Code of Conduct (HONcode) [48]. The QCSS is an instrument designed to offer scores on ownership, authorship, author qualification, purpose, attribution (references provided for requiring statements), interactivity, and currency of posting and revision. The sum of scores generates a final score summarized as excellent (80%-100%), very good (70%-79%), good (60%-69%), fair (50%-59%), or (poor 0%-50%). The HRWEF tool can be used by health professionals and patients to assess the appropriateness of websites. It consists of 30 items where each criterion is rated on a 3-point scale, scored as not applicable (score=0), disagree (score=1), or agree (score=2). It is divided into 7 main sections assessing the content, accuracy, author, currency, audience, navigation, and external links. An overall score was designated as excellent (90%-100%), adequate (75%-89%), or poor (0%-75%). Moreover, the HONcode certification validates and certifies the quality of the medical information provided on the internet.

Suitability of Information

The Suitability Assessment of Materials (SAM) tool [49] was used to assess the appropriateness of health information materials by considering characteristics such as content, graphics, literacy level, layout, typography, and cultural appropriateness of the websites. Each of the 22 items was rated as superior (rating +2), adequate (rating +1), not suitable (rating 0), or not applicable. Scores were summed to yield an overall percentage for the website reported as superior (70%-100%), adequate (40%-69%), or not suitable (0%-39%).

Readability

Readability tools were used to assess the difficulty of reading the written texts on the websites. The Flesh Kincaid test (F-K) [50], Simple Measure of Gobbledygook (SMOG) [51], and readability consensus based on 7 readability formulas (Flesch Reading Ease score, Gunning Fog, F-K, SMOG, the Coleman-Liau Index, Automated Readability Index, and Linsear Write Formula) [52] were used. The reviewers assessed the readability by selecting multiple written sections from each website and inserting it into a web-based readability calculator [52] that calculated F-K, SMOG, and readability consensus scores. In addition, as an item of the SAM tool, readability was also assessed using SMOG and rated as superior (grade 5 or lower), adequate (6th-8th grade level), or not suitable (grade 9 or higher). The Australian government recommends aiming for a lower than grade 8 reading level for health information [53-55], whereas the American Medical Association recommends education materials to be written at grade 6 reading level or lower [56].

Website Interactivity and Features

A validated interactivity scale was used as an individual consumers' perceptual assessment of websites in a previous study, which asked undergraduate business students ranging from age 19 to 40 years were asked to browse and rate websites using 15 items based on their personal experience [57]. The 15 items measuring active control (control over what users can do and see on the websites), two-way communication (ease of communication and offering feedback on the website), and synchronicity (website responsiveness to input and obtaining instantaneous information) were adopted for the purpose of this study.

A 3-point Likert scale was created to score each item as follows: agree (score=2), partially agree (score=1), or disagree (score=0), and an average score for all components was calculated. Interactivity scale was summarized as excellent (≥90%), adequate (70%-89%), or poor (≤69%).

Interactive aspects and features were also assessed by looking at whether the website was functional on a smartphone screen, had an associated app, addressed ethnicity, and included language options, paid features, search functions, games, videos, podcasts, chatbot, question and answer forum, quizzes, animation, a feedback form, slide shows, ratings, frequently asked questions section, recipes, read out loud options, navigation menu, social media links, acceptable page speed, webinars, or other.

Statistical Analysis

Interrater Reliability

Authors DJ and HC undertook interrater reliability (IRR) checking. A random 10% sample of all websites (n=7) were selected—the coding by DJ and HC were compared, and an IRR score was generated. Discrepancies were discussed until reviewers reached a consensus on their final ratings. Any disagreements were resolved by a third reviewer (ST).

IRR was calculated for the readability scores, quality of content scores, SAM, and the evaluation of information content using intraclass correlation coefficients (ICCs), with a high ICC value (maximum 1.0 indicating no variance in the scoring between different assessors, whereas ≥0.5 was moderate, ≥0.70 was good, and ≥0.80 indicated excellent reliability).

Software Used

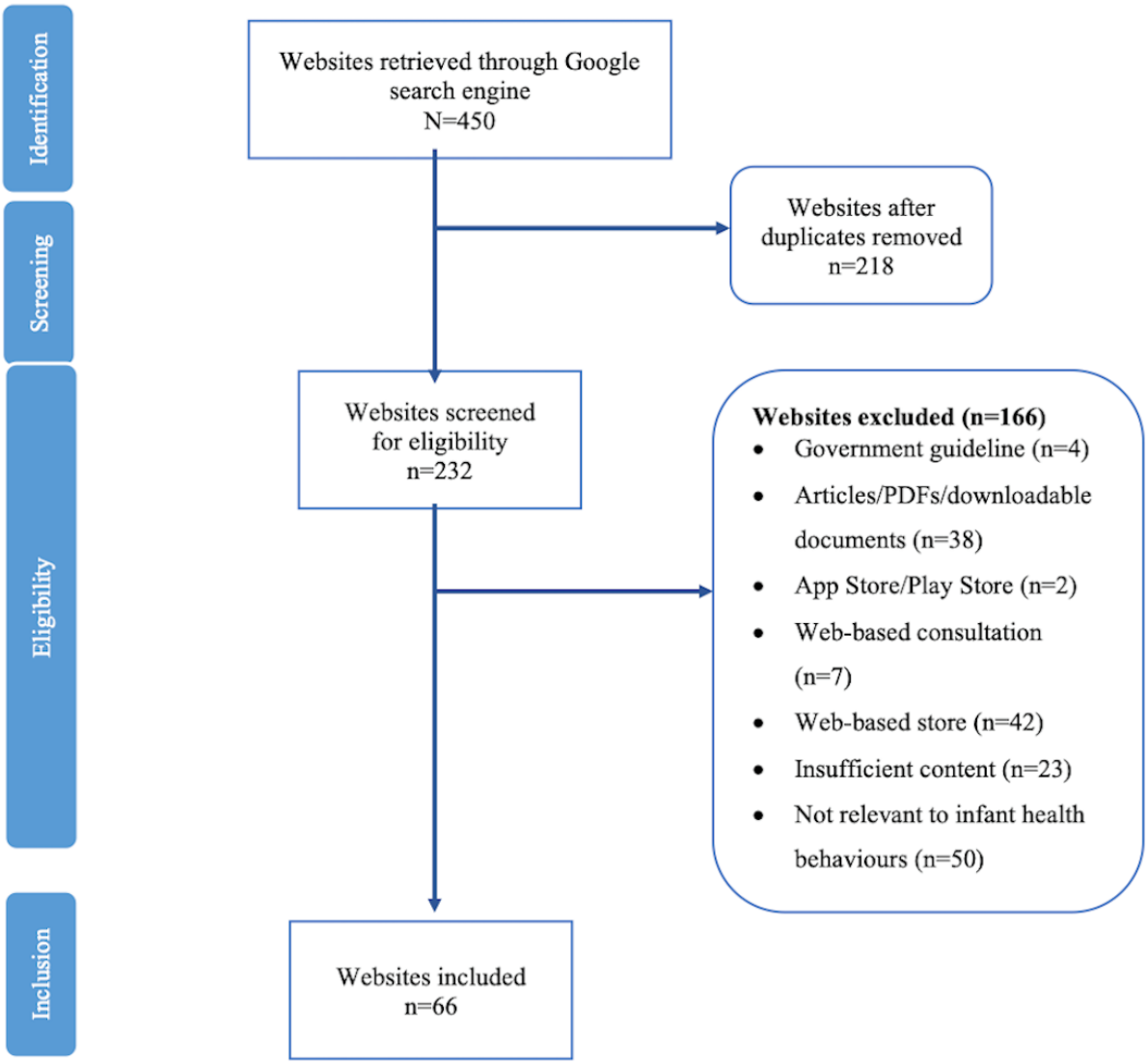
Data were transferred from REDCap to SPSS for MacBook (version 27.0; IBM Corp), where statistical analyses were performed. The ICC values calculated for content, HRWEF, QCSS, interactivity, and SAM were 0.5, 0.6, 0.6, 0.7, and 0.7, respectively, indicating a moderate to good level of consistency for the rating measurements. As the readability grades were calculated using computerized software, interrater consistency was not measured. The reviews discussed discrepancies by re-evaluating the websites together to ensure scoring consensus.

Results

Screening Process

As shown in Figure 1, a total of 450 global websites were reviewed between August 2021 and February 2022. The removal of 218 duplicate websites left a total of 232 unique websites. Of these, 66 websites met the inclusion criteria and were eligible to be evaluated. The remaining 166 websites were excluded, as 50 were not relevant to infant health behaviors; 38 were articles, PDF documents, or downloadable documents; 42 were a web-based shops; 23 had insufficient content; 4 were government guidelines; 7 offered web-based consultations, and 2 were infant-related apps on Google Play and App Store.

Figure 1. Diagram of website selection process.



Scope, Accuracy, and Depth of Information

Scope and Depth of Subtopics

The scope and depth of the information covered in the subtopics were predominantly poor, with 91% (60/66) of websites obtaining an overall score of ≤74%, whereas only 9% (6/66) of websites were rated as adequate.

The overall mean rating of all websites was poor (53%, SD 18%; IQR 40%-67%; Table 2). Expressing, feeding, and storage of expressed breast milk, preparing and feeding infant formula practices, and monitoring infant's progress topics had the lowest mean scores of 33%, 43%, and 49%, respectively. For information on infant sleep recommendations and bedtime practices, active play, screen time, and breastfeeding recommendations, correct advice was mostly reported as reflected by their mean scores of 73%, 70%, and 66%, respectively. Only 17% (10/58) of websites fully addressed infant feeding to appetite by encouraging responsiveness to infant hunger and satiety cues, not pressuring the baby to finish the bottle, feeding to appetite or baby-led feeding, avoiding bottle propping and bottle use in bed, and benefits of allowing infants to self-regulate their own appetite. Infant feeding to appetite was partially addressed by 59% (34/58) of the websites by highlighting a few of the above-mentioned points (Table 2 and Multimedia Appendix 3).

Table 2. The quantitative scope and depth of information based on Australian infant feeding and physical activity guidelines on all websites (N=66).

Topics addressed and websites[a]	Values, mean (SD)	Values, median
Breastfeeding		
Breastfeeding recommendations (n=58)	66 (21)	71
Physiology of breast milk and breastfeeding (n=55)	62 (21)	62.5
Monitoring infant's progress (n=56)	49 (26)	50
Breastfeeding, common problems, and their management		
Maternal factors affecting breastfeeding (n=54)	52 (26)	54
Infant factors affecting breastfeeding (n=54)	50 (26)	50
Expressing and storing breast milk		
Expressing, feeding, and storage of expressed breast milk (n=50)	33 (32)	30
Infant formula		
Preparing and feeding infant formula practices (n=58)	43 (19)	46
Introducing solids		
Solid introduction and foods and beverages not suitable for infants (n=58)	50 (23)	50
Infant activity		
Active play and screen time (n=52)	70 (30)	77.5
Infant sleep		
Cosleep recommendations (n=53)	61 (21)	62.5
Sleep recommendations and bedtime practices (n=54)	73 (29)	75
Overall content		
Overall scope and depth of information (n=66)	53 (18)	55

[a]Not all websites included information on all subtopics.

Subtopics Addressed

Subtopics that were most frequently correctly addressed on websites were recommendation to exclusively breastfeed till 6 months of age and continue breastfeeding with appropriate complementary food till 12 months of age and beyond (45/58, 77% of websites); natural patterns of breastfeeding 8 to 12 times over 24 hours (42/55, 76% of websites); postnatal breastfeeding advice to seek support from lactation consultants, midwives, or doctors (40/58, 68% of websites); tummy time recommendations (40/52, 76% of websites); and cosleeping in a separate cot but the same room as parents for the first 6- to 12-month recommendation (39/53, 73% of websites).

Subtopics Not Addressed

Subtopics that were most often not addressed on websites were supplemental requirements for infants on a vegan diet (30/58, 51% of websites); sterilization and proper use of hand pumps (30/50, 60% of websites); factors affecting initiation of lactation after birth (30/54, 55% of websites); importance and awareness of baby-friendly hospital initiative (27/58, 46% of websites); and correct water temperature for preparing infant formula and risk of infection from *Cronobacter sakazakii* bacteria (26/58, 44% of websites).

Subtopics Incorrectly Addressed

Subtopics that were most frequently incorrectly addressed were storage of freshly expressed, thawed, or used breast milk (22/50, 44% of websites); correct selection of infants' first foods (6/58, 10% of websites); and correct preparation of infant formula (5/58, 8% of websites).

Assessment of Website Quality

Using HRWEF

A majority of the websites attained an adequate rating (49/66, 74%) for the quality of the websites using the HRWEF tool. Although 20% (13/66) of the websites were rated as excellent, the remaining 3% (4/66) of websites received a poor scoring, 3 of which were commercial.

The overall HRWEF mean percentage score was 85% (SD 5.98%). The questions with the highest scores addressed the organization of the site, navigation, internal link, and the type of audience the author targeted. Conversely, the questions with the lowest scores were related to dates of publication and revision of content (Multimedia Appendix 3).

Using the QCSS

From the quality evaluation conducted using the QCSS tool, 8% (5/66) were rated as excellent, 21% (14/66) as very good, 32% (21/66) as good, 11% (7/66) as fair, and 29% (19/66) were rated as poor. The overall mean QCSS score was 60 (SD 18; Multimedia Appendix 3).

In comparison with websites that scored excellent, poorly rated websites failed to provide references, author qualifications, and

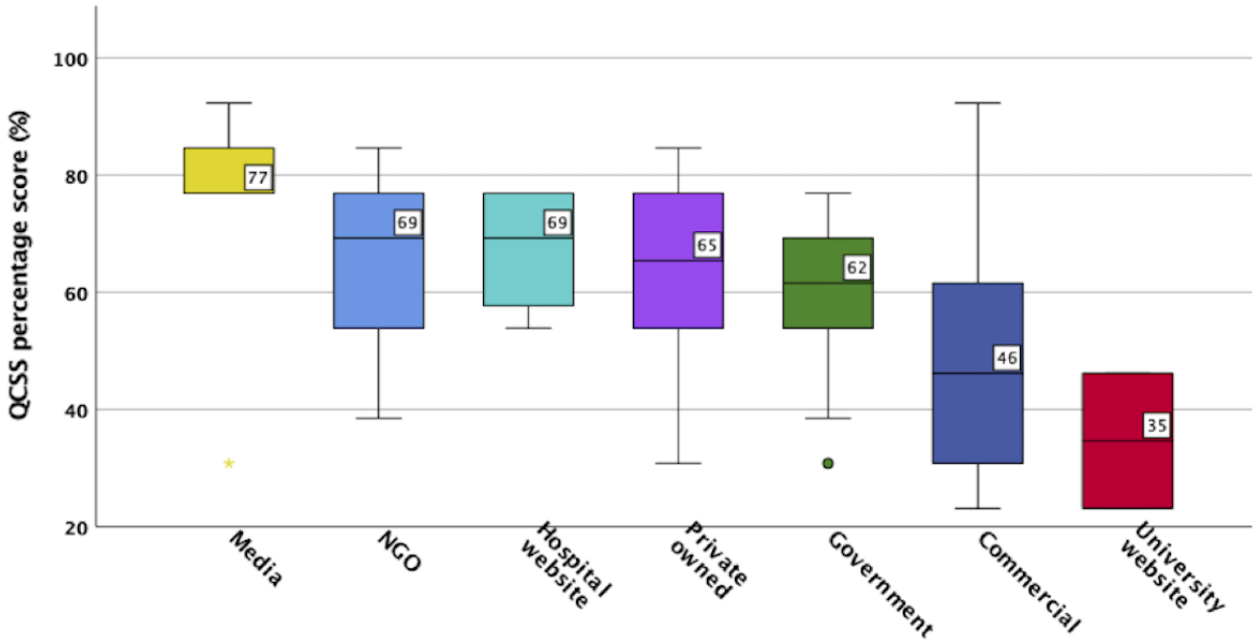
currency of content. A total of 32 websites stated that the author was a health care professional, whereas 30 websites clearly listed the name of the person supplying the information and author qualification. In addition, 31 websites had failed to display references for requiring statements. Only 3 websites presented the dates of original posting and revision.

A total of 9 websites stated they had acquired the HONcode certification demonstrating the intent of offering quality health information to meet ethical standards.

Quality (QCSS) by Organization

Figure 2 shows the quality of the websites by the type of organization as measured by QCSS. This ranged from poor to very good. Media-related sources received the highest median score of 77% (very good), followed by nongovernmental organizations, hospital websites, and privately owned and government websites, which received a score corresponding to good. Commercial websites had a mean score of 46% (fair), and university websites received the lowest median score of 35% (poor).

Figure 2. Quality Component Scoring System (QCSS) by type of organization. NGO: nongovernmental organizations.



Assessment of Suitability of Website Information by SAM

Overall, 3% (2/66 assessments) of websites were rated *superior* for suitability of health information, 82% (54/66 assessments) of websites as *adequate*, and 15% (10/66 assessments) of websites *not suitable*, as shown in Table 3 (Multimedia Appendix 3).

Very few websites were rated superior on literacy demand, such as writing style, context, and vocabulary used. Overall, 15% of websites provided culturally appropriate visual aids based on the consumers they were targeting. There were variations in the type of images used in the resources. Some images depicted different sex, race, color, religion, and age, whereas others targeted specific cultural groups such as Indian and Aboriginal and Torres Strait Islander. Only one of the sites addressed the cultural specificity of information relating to experience, language, or provision of examples to patients from diverse sociodemographic backgrounds.

Overall, 21% (14/66) of websites had the option to be translated into a language other than English, such as Arabic, Spanish, Hindi, or Bengali. Only 14% (9/66) presented information that addressed culture in texts or images. The culturally appropriate information varied between fasting tips and breastfeeding in Islam, Christianity or Judaism, fasting and pregnancy tips, multiracial graphics, and recipes.

Many websites provided a clear layout of information and easily understandable cover graphics that clearly portrayed the purpose of the material. Most topics were subdivided to improve readers' self-efficacy, which was rated as *adequate*; for instance, infant sleep was subdivided into quiet playtime, bedtime routine, safety sleep practices, and infant sleep recommendations.

Table 3. Website scores based on the Suitability Assessment of Materials (SAM) criteria.

	SAM scores (evaluations), n (%)			
	Not suitable	Adequate	Superior	Not applicable
Content				
Purpose is evident	1 (1)	30 (45)	35 (53)	—[a]
Content about behaviors	2 (3)	59 (89)	5 (8)	—
Limited to essential information	4 (6)	40 (61)	22 (33)	—
Summary and review	46 (70)	16 (24)	3 (5)	1 (1)
Literacy demand				
Reading grade level	31 (47)	31 (47)	3 (5)	1 (1)
Writing style with active voice	5 (8)	55 (83)	6 (9)	—
Vocabulary uses common words	5 (8)	56 (84)	5 (8)	—
Context given first	11 (17)	47 (71)	8 (12)	—
Headers or topic captions	1 (1)	21 (32)	44 (67)	—
Graphics				
Purposeful cover graphic	9 (14)	26 (39)	30 (46)	1 (1)
Appropriate type of illustrations	9 (14)	22 (33)	6 (9)	29 (44)
Relevance of illustrations	11 (17)	19 (29)	8 (12)	28 (42)
Lists, tables, graphs, and charts explained	7 (11)	45 (68)	5 (8)	9 (14)
Captions used for graphics	9 (14)	41 (62)	11 (17)	5 (8)
Layout and typography				
Layout factors	6 (10)	58 (88)	2 (3)	—
Typography	2 (3)	59 (89)	5 (8)	—
Subheadings used	0 (0)	42 (64)	24 (36)	—
Learning, stimulation, and motivation				
Interaction with readers used	22 (33)	44 (67)	0 (0)	—
Modeled and specific behaviors	16 (24)	45 (68)	5 (8)	—
Self-efficacious tasks and behaviors	5 (8)	57 (86)	4 (6)	—
Cultural appropriateness				
Cultural match	0 (0)	1 (1)	0 (0)	65 (99)
Cultural image and examples	0 (0)	8 (12)	2 (3)	56 (85)

[a]Not included in the overall score.

Assessment of Website Readability

Very few websites met the Australian Federal government's recommended level for written health information of lower than a grade 8 reading level: 29% (19/66) of websites (SMOG), 20% (13/66) of websites (F-K web-based tool), and 12% (8/66) of websites (consensus tool). In 2 of the websites assessed for interreliability, the readability scores ranged from 8 to 11 and 7 to 14 between the researchers depending on the varying content selected for assessment (Multimedia Appendix 4).

The median readability grades were 8.5 (IQR 7-10), 9 (IQR 8-11), and 10 (IQR 8-11) using the SMOG formula, web-based F-K calculator, and consensus calculator, respectively. There was a good correlation among reading grade scores across the readability measures (P<.001; 2-tailed).

Assessment of Website Interactivity and Features

Table 4 presents the results of the websites interactivity scores in terms of active control, two-way communication, and synchronicity.

Table 4. Interactivity scores of websites (N=66).

	Agree, n (%)
Active control	
I felt that I had a lot of control over my visiting experiences at this website	17 (25)
While I was on the website, I could choose freely what I wanted to see	23 (34)
While surfing the website, I had full control over what I can do on the site	21 (31)
While surfing the website, my actions decided the kind of experiences I got	24 (36)
Two-way communication	
The website is effective in gathering visitors' feedback	6 (9)
This website facilitates two-way communication between the visitors and the site	7 (10)
It is easy to offer feedback to the website	13 (19)
The website makes me feel it wants to listen to its visitors	9 (13)
The website encourages visitors to talk back	8 (12)
The website gives visitors the opportunity to talk back	18 (27)
Synchronicity	
The website processed my input very quickly	16 (24)
Getting information from the website is very fast	15 (22)
I was able to obtain the information I want without any delay	10 (15)
When I clicked on the links, I felt I was getting instantaneous information	26 (39)
The website was very fast in responding to my requests	18 (27)

The overall interactivity of websites was predominantly poor (49/66, 74%). The remaining 26% (17/66) of websites were marginally adequate with no websites obtaining an excellent rating. More than half the websites acquired an incomplete score for active control resulting from slow loading web pages and the inability of site search engines to return relevant results effectively. Very few websites encouraged visitors to talk back or facilitated 2-way communication between the visitors and the site. As for synchronicity of the sites, approximately one-fourth of websites received a full score (agree) for their ability to process input and respond to requests promptly (Multimedia Appendix 4).

The most common features found on websites were social media links (61/66, 92%), frequently asked questions (48/66, 73%), videos (44/66, 67%), and recipes (35/66, 53%), whereas language options, webinars, question and answer forums, chatbots, read out loud function, slide shows, animation, and games were less common. Moreover, 80% (53/66) of websites had additional features such as text font size options, tools (eg, ovulation calculator, pregnancy calculator, and parenting tools), download and print page content option, and YouTube accounts. Overall, 47% (31/66) of websites had associated apps on Google Play and Apple Store. Log-in options for personalized health information were presented on 40% (26/66) of the websites.

Discussion

Principal Findings

In this review, we systematically assessed 66 websites that reported health information related to infant nutrition, active play, screen time, or sleep behaviors. This review extends on the existing 2015 assessment by providing 2 main conceptual contributions. First, it covers the quality, content, suitability, readability, and comprehensibility of web-based infant health information at a global level. Second, it assesses the interactivity, features, and cultural considerations of webpages. In this section, we discuss the principal findings, comparison to prior work, implications for future practice, and then outline the strengths and limitations of this review.

This study found that the information content of the websites was overall poor in terms of scope and depth of information, which was similar to the findings of the previous 2015 assessment. Approximately one-third of the websites reported different advice on storage of expressed breast milk; for instance, "freshly expressed breast milk can be stored safely in the refrigerator for up to five days." This information was contrary to the Australian guidelines on infant feeding that stated storage of expressed breast milk should not exceed 72 hours in the fridge [42]. This was due to the development of some websites in other countries, such as America or Europe, where guidelines differ from that of Australia [58,59].

The quality of websites was generally adequate when evaluated by HRWEF and mostly ranged from poor to good when rated by the QCSS tool. These findings are consistent with those from previous studies [60-63], which evaluated a range of health information available on the web using similar tools. This study also highlights that the quality of websites in terms of ownership, authorship, author qualification, purpose, referencing statements, and currency of information was the highest among media, nongovernmental organizations, and hospital websites and the lowest among university websites. This is an important finding,

given that parents view university sites as a high-quality, reliable, and credible source of information [64]. It is vital that the health information on websites is continuously updated to meet the latest guidelines with relevant currency, authorship, qualification, and supporting attribution statements. This in turn will provide readers with the clarity they need to assess the quality of web-based health information and identify reputable websites. Moreover, websites with outdated information and no supporting statements can mislead readers, resulting in adverse health consequences [65-67].

Using the SAM tool, we found that website information on infant health behaviors was generally adequate. This finding is consistent with that of other studies on the SAM [68-70]. Despite the overall adequate suitability ratings of information on the selected websites, there were marked limitations in terms of cultural appropriateness, literacy demand, and illustrations. This highlights the issue that web-based infant health information rarely considered the needs of people from non–English-speaking ethnic groups and how they may interpret or apply the health information. This is unfortunate, given the ubiquitous nature of the internet. Given the increase in cultural diversity within Australia and abroad, it is important to consider cultural appropriateness of information and provide culturally and ethnically diverse consumers the capacity to access, understand, and use health information to make well-informed health decisions [71-73].

In regard to readability, this study highlighted that most websites were at readability levels beyond the ideal level of lower than grade 8. This finding was also reflected in previous studies [74-76]. Notably, a difference of 4 to 6 grades was observed between the interreliability readability grades of 2 websites scored by the 2 researchers DJ and HC. This reflected the inconsistent readability levels across various webpages within a website. Readability and health literacy play an integral part in information accessibility and usability [77]. The readability formulas are based on the number of words, sentence length, and number of syllables per word. Therefore, using simpler words, shorter sentences, pictures, videos, and co-design methods as per the Australian Commission for Safety and Quality in health care are important considerations for writing health information for consumers [78].

Furthermore, readability levels of web-based health information should be tested for consistency and presented in an easy-to-read format providing access to people with low health literacy.

The overall interactivity of website functions was poor. In addition, interactive features were mostly common among media and commercial websites and least common among government websites. According to the World Health Organization, various provision methods of providing health information are important to increase accessibility and achieve positive health outcomes [79]. Multiple patient-focused interventions have reported that various health information formats, such as videos, audios, and infographics, have contributed to the improvement in parental knowledge, satisfaction, and health outcomes [80-82]. Hence, it is crucial for credible websites such as government owned to make the wealth of information available on the web interactive and accessible to increase consumer engagement and use of reliable sources.

Comparison With Prior Work

In comparison with the 2015 review, 55 new websites were assessed in this review, whereas 11 were common across both studies. It is important to note that the first 30 websites generated from every search term used resulted in a range of global websites for consumers to access. Hence, it is vital for education to be provided to parents of young children and health professionals that will enable them to determine the quality and credibility of web-based health information as accuracy is critical, especially in the first 1000 days of life [83,84].

Furthermore, several websites from the 2015 review were excluded for reasons such as their web-based content had been removed or they no longer exist. A potential reason may be due to the cost and maintenance of websites that were developed with limited funding. A systematic review of factors that influence eHealth reported that ongoing maintenance costs were barriers for several studies [85]. Another reason may be due to the evolution of websites over time and the inability of static websites to enable and host new features [30].

To replicate and compare results from the 2015 assessment, the same validated tools were used to assess the quality, suitability, and readability of infant health websites. Interestingly, the results of the new eligible websites did not greatly differ from the results of the 2015 systematic assessment [25]. There was a slight improvement in the quality of websites from poor to adequate or good measured by HRWEF and QCSS since 2015. However, the suitability rating of websites in both assessments measured by SAM was the same. Similarly, the readability of written health information in the majority of the websites did not meet government recommendation in both assessments as well. Although websites have evolved since 2015, minimal to no improvements have been identified in this review. One reason could be due to the lack of user involvement in the website design. A recent systematic review reported that a considerable number of studies raised concerns that involving users in technology design can be fairly demanding and requires time and effort [86]. Another reason may be the lack of use of validated tools such as the ones used in this study to ensure optimized quality during the development of websites.

Identifying cultural considerations, interactivity, and features of web pages is a value added to this study. We found that very few websites addressed culture or had interactive features such as multilingual options, chatbots, or read out loud functions. With the rise in immigrants from culturally and linguistically diverse communities [87], more and more people are facing access barriers to health information and eHealth services due to the lack of language support and culturally appropriate health information through the internet [88-90]. This demonstrates the need to engage end users throughout the web development to ensure high-quality outcomes and meet consumers' needs and expectations.

Implications for Practice

With the wide spread of internet use and wealth of information available on the web [91], it is imperative for parents and health

professionals to be guided to the optimal and most accurate sources of information. Websites should be screened for authorship, ownership, information date, and HONcode certification before use. There is a need for web developers to bear in mind their end users. One way to overcome this challenge is by involving consumers in website development through co-design workshops [92]. This will ensure developers get a good understanding of end users' access requirements, literacy demands, preference for alternative presentation formats of information, and cultural considerations. An establishment of a regulatory body is also recommended to ensure that newly developed websites are built on validated tools and that all websites comply with government guidelines standards.

Strengths and Limitations

To the best of our knowledge, this is the first study to evaluate website interactivity, features, and cultural considerations for web-based infant health information. This review provides a comprehensive global overview of the available web-based information about infant health behaviors and identifies ways for improvement.

Although this study adhered to a rigorous systematic search process, there were several limitations. First, most websites included in this evaluation differed from that which was evaluated in the 2015 review. Hence, this demonstrates the dynamic nature and constant change of the internet. Thus, the website search during this review reflected a period that could potentially change. In addition, the website search was conducted in the English language using Google. Although Google is a highly used search engine, we acknowledge that some international users have access restraints [93]. Therefore, the results may not have identified websites present on other search engines or written in other languages. Furthermore, deleting cookies and search histories was intended to reduce unknown bias in the search strategy. However, it is acknowledged that the likelihood of most web users doing this is unlikely, and their searches might identify sites of which we were unaware.

Another limitation is that the assessment criteria used Australian guidelines. Therefore, there was a potential for websites following non-Australian guidelines to obtain an incorrect score on a few subtopics. Moreover, the interactivity scale used was originally meant to capture consumers' perceptual assessment of websites. However, due to the lack of published validated tools used to measure interactivity of websites, the interactivity scale was adopted for the purpose of this study. Thus, the interactivity scale used was a subjective measurement based on the researcher's experience on the websites. Nevertheless, the tools used offer a standardized way to best capture the quality and interactivity of web-based information.

Conclusions

As more parents seek web-based guidance on infant health behaviors globally, there remains a significant concern on the quality, readability, interactivity, and accessibility of websites promoting health behaviors during infancy. This systematic assessment revealed that there is a need for researchers and health care providers to leverage innovative web-based platforms to provide culturally responsive evidence-based information accessible to those with limited English proficiency. Furthermore, a focus is needed on continuously updating existing health websites in addition to recommending an establishment of a regulatory body to ensure compliance with government standards. Moreover, the development of new eHealth technology should be based on validated tools to ensure the optimal quality of websites.

Acknowledgments

This work is a component of PhD degree of DJ at the University of Sydney.

Authors' Contributions

DJ and ST designed the study and the main conceptual ideas. DJ undertook the website search and evaluation. DJ and HC undertook the interrater reliability testing. DJ undertook statistical analysis. ST, LMW, LB, and CR cross-checked uncertain websites until consensus was reached. DJ wrote the first draft of the article; all authors critically revised the paper and provided comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Website evaluation tools.
[DOCX File , 31 KB - jmir_v24i10e38641_app1.docx]

Multimedia Appendix 2
Protocol.
[DOCX File , 38 KB - jmir_v24i10e38641_app2.docx]

Multimedia Appendix 3

Summary score for all websites.
[DOCX File , 24 KB - jmir_v24i10e38641_app3.docx]

Multimedia Appendix 4
Supplementary tables and graphs.
[DOCX File , 71 KB - jmir_v24i10e38641_app4.docx]

References

1. Poushter J. Internet access growing worldwide but remains higher in advanced economies. Pew Research Center. 2016 Feb 22. URL: https://www.pewresearch.org/global/2016/02/22/internet-access-growing-worldwide-but-remains-higher-in-advanced-economies/ [accessed 2022-03-28]
2. Roser M, Ritchie H, Ortiz-Ospina E. Internet. Our World in Data. 2015. URL: https://ourworldindata.org/internet [accessed 2022-03-28]
3. World Internet Users and 2022 Population Stats. Internet World Stats. URL: https://www.internetworldstats.com/stats.htm [accessed 2022-03-28]
4. Mobile Search Engine Host Market Share Worldwide. StatCounter Global Stats. URL: https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet [accessed 2022-03-28]
5. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. AMIA Annu Symp Proc 2003;2003:225-229 [FREE Full text] [Medline: 14728167]
6. Schneider JA, Holland CP. eHealth search patterns: a comparison of private and public health care markets using online panel data. J Med Internet Res 2017 Apr 13;19(4):e117 [FREE Full text] [doi: 10.2196/jmir.6739] [Medline: 28408362]
7. eHealth information search behaviour: Is there a difference between private and public health care markets? Research Features. 2020 Aug 12. URL: https://researchfeatures.com/ehealth-information-search-behaviour-difference-between-private-and-public-health-care-markets/ [accessed 2022-03-28]
8. Choi H, Ahn S. Classifications, changes, and challenges of online health information seekers during COVID-19. Int J Environ Res Public Health 2021 Sep 09;18(18):9495 [FREE Full text] [doi: 10.3390/ijerph18189495] [Medline: 34574422]
9. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802 [FREE Full text] [doi: 10.2196/29802] [Medline: 34043526]
10. Homeniuk R, Collins C. How COVID-19 has affected general practice consultations and income: general practitioner cross-sectional population survey evidence from Ireland. BMJ Open 2021 Apr 08;11(4):e044685. [doi: 10.1136/bmjopen-2020-044685]
11. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857 [FREE Full text] [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
12. 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]
13. Bouche G, Migeot V. Parental use of the Internet to seek health information and primary care utilisation for their child: a cross-sectional study. BMC Public Health 2008 Aug 28;8:300 [FREE Full text] [doi: 10.1186/1471-2458-8-300] [Medline: 18755029]
14. Lupton D, Maslen S. How women use digital technologies for health: qualitative interview and focus group study. J Med Internet Res 2019 Jan 25;21(1):e11481 [FREE Full text] [doi: 10.2196/11481] [Medline: 30681963]
15. Nievas-Soriano BJ, Castro-Luna GM, García-Duarte S, González-López MD, Parrón-Carreño T. Profile of the users and the most visited topics of a pediatric eHealth website. Int J Environ Res Public Health 2021 Oct 26;18(21):11248 [FREE Full text] [doi: 10.3390/ijerph182111248] [Medline: 34769765]
16. Mahony C, Heavin C, Sammon D. Health-related information resource use: an exploratory study of expectant mothers. J Decis Syst 2022;31(3):234-257 [FREE Full text] [doi: 10.1080/12460125.2021.1872893]
17. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 01;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
18. Dexter B, Frank S, Seguin L. How and where parents of infants and young children want to receive nutrition information. Can J Diet Pract Res 2016 Dec;77(4):203-205. [doi: 10.3148/cjdpr-2016-020] [Medline: 27763768]
19. Alianmoghaddam N, Phibbs S, Benn C. "I did a lot of Googling": a qualitative study of exclusive breastfeeding support through social media. Women Birth 2019 Apr;32(2):147-156. [doi: 10.1016/j.wombi.2018.05.008] [Medline: 29921552]
20. Slomian J, Bruyère O, Reginster JY, Emonts P. The internet as a source of information used by women after childbirth to meet their need for information: a Web-based survey. Midwifery 2017 May;48:46-52. [doi: 10.1016/j.midw.2017.03.005] [Medline: 28324809]

21. Asiodu IV, Waters CM, Dailey DE, Lee KA, Lyndon A. Breastfeeding and use of social media among first-time African American mothers. J Obstet Gynecol Neonatal Nurs 2015;44(2):268-278 [FREE Full text] [doi: 10.1111/1552-6909.12552] [Medline: 25712127]
22. Wright AL, Jack SM, Ballantyne M, Gabel C, Bomberry R, Wahoush O. Indigenous mothers' experiences of using primary care in Hamilton, Ontario, for their infants. Int J Qual Stud Health Well-being 2019 Dec;14(1):1600940 [FREE Full text] [doi: 10.1080/17482631.2019.1600940] [Medline: 31033431]
23. Rhodes A. Sources of child health information: what parents use and trust. Royal Children's Hospital National Child Health Poll. Melbourne, Australia: The Royal Children's Hospital; 2016. URL: https://www.rchpoll.org.au/polls/sources-of-child-health-information-what-parents-use-and-trust/ [accessed 2022-03-28]
24. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]
25. Taki S, Campbell KJ, Russell CG, Elliott R, Laws R, Denney-Wilson E. Infant feeding websites and apps: a systematic assessment of quality and content. Interact J Med Res 2015 Sep 29;4(3):e18 [FREE Full text] [doi: 10.2196/ijmr.4323] [Medline: 26420339]
26. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
27. Son HM, Je M, Sohn YS. Quality evaluation of online health information related to young child. Child Health Nurs Res 2018 Jan 31;24(1):91-100 [FREE Full text] [doi: 10.4094/chnr.2018.24.1.91]
28. Dol J, Richardson B, Boates T, Campbell-Yeo M. Learning to parent from Google? Evaluation of available online health evidence for parents of preterm infants requiring neonatal intensive care. Health Informatics J 2019 Dec;25(4):1265-1277 [FREE Full text] [doi: 10.1177/1460458217752564] [Medline: 29361867]
29. Król K. Evolution of online mapping: from Web 1.0 to Web 6.0. Geomat Landmanag Landsc 2020;1:33-51 [FREE Full text] [doi: 10.15576/gll/2020.1.33]
30. Web 1.0, web 2.0 and web 3.0 with their difference. GeeksforGeeks. 2022. URL: https://www.geeksforgeeks.org/web-1-0-web-2-0-and-web-3-0-with-their-difference/ [accessed 2022-03-28]
31. Cisco Annual Internet Report (2018–2023) White Paper. Cisco. 2022. URL: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html [accessed 2022-03-28]
32. International bandwidth in the Asia-Pacific region in 2020, by country or territory. Statista. 2020. URL: https://www.statista.com/statistics/1293199/apac-international-bandwidth-by-country/ [accessed 2022-03-28]
33. Armstrong M, Richter F. Infographic: How many websites are there? Statista. 2021. URL: https://www.statista.com/chart/19058/number-of-websites-online/ [accessed 2022-03-28]
34. Harrison C, Britt H, Wong C, Henderson J. Patient use of the internet for health information. Aust Fam Physician 2014 Dec;43(12):875-877 [FREE Full text] [Medline: 25705739]
35. Digital Health. Australian Institute of Health and Welfare. 2022 Jul 7. URL: https://www.aihw.gov.au/reports/australias-health/digital-health [accessed 2022-03-28]
36. Sawyer R. The Impact of New Social Media on Intercultural Adaptation. Senior Honors Projects. 2011 May. URL: https://digitalcommons.uri.edu/srhonorsprog/242/ [accessed 2022-03-29]
37. Morrison C, Doherty G. Analyzing engagement in a Web-based intervention platform through visualizing log-data. J Med Internet Res 2014 Nov 13;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
38. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019 Sep 30;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
39. Adamopoulou E, Moussiades L. An overview of chatbot technology. In: Proceedings of the 16th International Conference on Artificial Intelligence Applications and Innovations. 2020 Presented at: AIAI '20; June 5–7, 2020; Neos Marmaras, Greece p. 373-383. [doi: 10.1007/978-3-030-49186-4_31]
40. Wong J, Foussat AC, Ting S, Acerbi E, van Elburg RM, Mei Chien C. A chatbot to engage parents of preterm and term infants on parental stress, parental sleep, and infant feeding: usability and feasibility study. JMIR Pediatr Parent 2021 Oct 26;4(4):e30169 [FREE Full text] [doi: 10.2196/30169] [Medline: 34544679]
41. Jacobsen M. How Far Down the Search Engine Results Page Will Most People Go? Leverage Marketing. 2017. URL: https://www.theleverageway.com/blog/how-far-down-the-search-engine-results-page-will-most-people-go/ [accessed 2022-04-03]
42. Infant Feeding Guidelines: Information for Health Workers. National Health and Medical Research Council. Canberra, Australia: Australian Government; 2012. URL: https://www.eatforhealth.gov.au/sites/default/files/files/the_guidelines/n56_infant_feeding_guidelines.pdf [accessed 2022-03-29]
43. Australian 24-Hour Movement Guidelines for the Early Years (Birth to 5 Years). Department of Health, Australian Government. Canberra, Australia: Australian Government; 2010. URL: https://www.health.gov.au/sites/default/files/documents/2021/05/24-hour-movement-guidelines-birth-to-5-years-brochure.pdf [accessed 2022-09-20]

44. For infants, toddlers and preschoolers (birth to 5 years). Department of Health and Aged Care, Australian Government. 2015. URL: https://www.health.gov.au/health-topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians/for-infants-toddlers-and-preschoolers-birth-to-5-years [accessed 2022-03-29]
45. Pealer LN, Dorman SM. Evaluating health-related Web sites. J Sch Health 1997 Aug;67(6):232-235. [doi: 10.1111/j.1746-1561.1997.tb06311.x] [Medline: 9285869]
46. Ghezzi P, Chumber S, Brabazon T. Educating medical students to evaluate the quality of health information on the web. In: Floridi L, Illari P, editors. The Philosophy of Information Quality. Cham, Switzerland: Springer; 2014:183-199.
47. Martins EN, Morse LS. Evaluation of Internet websites about retinopathy of prematurity patient education. Br J Ophthalmol 2005 May;89(5):565-568 [FREE Full text] [doi: 10.1136/bjo.2004.055111] [Medline: 15834086]
48. Methodology. HONcode. URL: https://protect-au.mimecast.com/s/OFi7CROND2urv1q5xFOtFjF?domain=hon.ch [accessed 2022-04-03]
49. Doak L, Doak C, Root J. Suitability Assessment of Materials For Evaluation of Health-Related Information for Adults. Aspirius Library. 2008. URL: https://protect-au.mimecast.com/s/LM_uCVARKgClx6ogYHQ1YDU?domain=aspiruslibrary.org [accessed 2022-03-29]
50. Kincaid JP, Fishburne Jr RP, Rogers RL, Chissom BS. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Institute for Simulation and Training, University of Central Florida. 1975 Jan 1. URL: https://stars.library.ucf.edu/istlibrary/56 [accessed 2022-03-29]
51. McLaughlin HG. SMOG grading - a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
52. Automatic Readability Checker - (Our Free Text Readability Consensus Calculator). Readability Formulas. URL: https://readabilityformulas.com/free-readability-formula-tests.php [accessed 2022-04-03]
53. Style manual. Australian Government Coat of Arms. URL: https://www.stylemanual.gov.au/accessible-and-inclusive-content/literacy-and-access [accessed 2022-04-03]
54. National Report on schooling in Australia 2009. Australian Curriculum, Assessment and Reporting Authority. 2011. URL: https://docs.acara.edu.au/resources/National_Report_on_Schooling_in_Australia_2009_live.pdf [accessed 2022-04-03]
55. Health literacy: taking action to improve safety and quality. Australian Commission on Safety and Quality in Health Care. Sydney, Australia: ACSQHC; 2014 Aug. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/Health-Literacy-Taking-action-to-improve-safety-and-quality.pdf [accessed 2022-04-03]
56. Grabeel KL, Russomanno J, Oelschlegel S, Tester E, Heidel RE. Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. J Med Libr Assoc 2018 Jan;106(1):38-45 [FREE Full text] [doi: 10.5195/jmla.2018.262] [Medline: 29339932]
57. Liu Y. Developing a scale to measure the interactivity of websites. J Advert Res 2003 Jun 01;43(2):207-216. [doi: 10.1017/S0021849903030204]
58. Proper storage and preparation of breast milk. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/breastfeeding/recommendations/handling_breastmilk.htm [accessed 2022-04-03]
59. Food for infants and young children. Food Safety. URL: https://ec.europa.eu/food/safety/labelling-and-nutrition/specific-groups/food-infants-and-young-children_en [accessed 2022-04-03]
60. Lobato MZ, Martins AC. Quality assessment of online information on Brazilian domain sites related to dental implants. Brazilian J Implant Health Sci 2020;2(7):21-22 [FREE Full text] [doi: 10.36557/26748169]
61. Nasser S, Mullan J, Bajorek B. Assessing the quality, suitability and readability of internet-based health information about warfarin for patients. Australas Med J 2012;5(3):194-203 [FREE Full text] [doi: 10.4066/AMJ.2012862] [Medline: 22952566]
62. Floridi L, Illari P. The Philosophy of Information Quality. Cham, Switzerland: Springer International Publishing; 2014.
63. Chumber S, Huber J, Ghezzi P. A methodology to analyze the quality of health information on the Internet: the example of diabetic neuropathy. Diabetes Educ 2015 Feb;41(1):95-105. [doi: 10.1177/0145721714560772] [Medline: 25480397]
64. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
65. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
66. Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. J Med Internet Res 2008 Jun 30;10(2):e17 [FREE Full text] [doi: 10.2196/jmir.986] [Medline: 18653441]
67. Kim R, Kim HJ, Jeon B. The good, the bad, and the ugly of medical information on the Internet. Mov Disord 2018 May;33(5):754-757. [doi: 10.1002/mds.27324] [Medline: 29436748]
68. Martin CA, Khan S, Lee R, Do AT, Sridhar J, Crowell EL, et al. Readability and suitability of online patient education materials for glaucoma. Ophthalmol Glaucoma (forthcoming) 2022 Mar 14:S2589-4196(22)00051-5. [doi: 10.1016/j.ogla.2022.03.004] [Medline: 35301989]

69. Doubleday AR, Novin S, Long KL, Schneider DF, Sippel RS, Pitt SC. Online information for treatment for low-risk thyroid cancer: assessment of timeliness, content, quality, and readability. J Cancer Educ 2021 Aug;36(4):850-857. [doi: 10.1007/s13187-020-01713-5] [Medline: 32108292]
70. Guan Y, Maloney KA, Roter DL, Pollin TI. Evaluation of the informational content, readability and comprehensibility of online health information on monogenic diabetes. J Genet Couns 2018 Jun;27(3):608-615 [FREE Full text] [doi: 10.1007/s10897-017-0155-y] [Medline: 28951986]
71. Institute of Medicine (US) Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington, DC, USA: National Academies Press (US); 2004.
72. Hughson JA, Marshall F, Daly JO, Woodward-Kron R, Hajek J, Story D. Health professionals' views on health literacy issues for culturally and linguistically diverse women in maternity care: barriers, enablers and the need for an integrated approach. Aust Health Rev 2018 Feb;42(1):10-20. [doi: 10.1071/AH17067] [Medline: 29081348]
73. Peterson EB, Chou WY, Kelley DE, Hesse B. Trust in national health information sources in the United States: comparing predictors and levels of trust across three health domains. Transl Behav Med 2020 Oct 08;10(4):978-988 [FREE Full text] [doi: 10.1093/tbm/ibz066] [Medline: 31116400]
74. Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual 2018;33(5):487-492. [doi: 10.1177/1062860617751639] [Medline: 29345143]
75. Parr O, Dunmall K. An evaluation of online information available for women with breast implants aged 47-73 who have been invited to attend the NHS Breast Screening Programme. Radiography (Lond) 2018 Nov;24(4):315-327. [doi: 10.1016/j.radi.2018.03.008] [Medline: 30292500]
76. Fefer M, Lamb CC, Shen AH, Clardy P, Muralidhar V, Devlin PM, et al. Multilingual analysis of the quality and readability of online health information on the adverse effects of breast cancer treatments. JAMA Surg 2020 Aug 01;155(8):781-784 [FREE Full text] [doi: 10.1001/jamasurg.2020.1668] [Medline: 32520317]
77. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
78. Manser G. Writing Health Information for Consumers. Canberra Health Literacy. 2022. URL: https://cbrhl.org.au/health-services-providers/writing-patient-information/ [accessed 2022-04-04]
79. Coulter A, Ellins J. Patient-Focused Interventions: A Review of the Evidence. Quest for Quality and Improved Performance. Oxford, UK: Picker Institute Health Foundation; 2006. URL: https://www.health.org.uk/sites/default/files/PatientFocusedInterventions_ReviewOfTheEvidence.pdf [accessed 2022-04-04]
80. Bandyopadhyay M, Stanzel K, Hammarberg K, Hickey M, Fisher J. Accessibility of Web-based health information for women in midlife from culturally and linguistically diverse backgrounds or with low health literacy. Aust N Z J Public Health 2022 Jun;46(3):269-274. [doi: 10.1111/1753-6405.13192] [Medline: 34939717]
81. Romantika IW, Lusmilasari L, Prabandari YS, Syahrul S. Application of video-based health education in improving mother's knowledge and attitudes about behavioral problems among preschool children. Enfermería Clínica 2020 Mar;30:172-176. [doi: 10.1016/j.enfcli.2019.07.071]
82. Shah N, Mathur VP, Kathuria V, Gupta T. Effectiveness of an educational video in improving oral health knowledge in a hospital setting. Indian J Dent 2016;7(2):70-75 [FREE Full text] [doi: 10.4103/0975-962X.184646] [Medline: 27433049]
83. Mihrshahi S, Jawad D, Richards L, Hunter KE, Ekambareshwar M, Seidler AL, et al. A review of registered randomized controlled trials for the prevention of obesity in infancy. Int J Environ Res Public Health 2021 Mar 02;18(5):2444 [FREE Full text] [doi: 10.3390/ijerph18052444] [Medline: 33801485]
84. Pietrobelli A, Agosti M, MeNu Group. Nutrition in the first 1000 days: ten practices to minimize obesity emerging from published science. Int J Environ Res Public Health 2017 Dec 01;14(12):1491 [FREE Full text] [doi: 10.3390/ijerph14121491] [Medline: 29194402]
85. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
86. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
87. World Migration Report 2020. International Organization for Migration (IOM). New York, NY, USA: United Nations; 2019. URL: https://publications.iom.int/system/files/pdf/wmr_2020.pdf [accessed 2022-03-26]
88. Pandey M, Maina G, Amoyaw J, Li Y, Kamrul R, Michaels C, et al. Impacts of English language proficiency on healthcare access, use, and outcomes among immigrants: a qualitative study. BMC Health Serv Res 2021 Jul 26;21(1):741 [FREE Full text] [doi: 10.1186/s12913-021-06750-4] [Medline: 34311712]
89. Chae S, Lee YJ, Han HR. Sources of health information, technology access, and use among non-english-speaking immigrant women: descriptive correlational study. J Med Internet Res 2021 Oct 29;23(10):e29155 [FREE Full text] [doi: 10.2196/29155] [Medline: 34714249]

90. Baker AE, Procter NG, Ferguson MS. Engaging with culturally and linguistically diverse communities to reduce the impact of depression and anxiety: a narrative review. Health Soc Care Community 2016 Jul;24(4):386-398. [doi: 10.1111/hsc.12241] [Medline: 25939369]
91. Coughler C, Burke SM, Cardy JO. Analysis of the quality of online resources for parents of children who are late to talk. Autism Dev Lang Impairments 2020 Apr 06;5:239694152091794 [FREE Full text] [doi: 10.1177/2396941520917940]
92. Co-Design Position Statement. Health Care Consumers' Association Executive Committee. 2020. URL: https://www.hcca.org.au/wp-content/uploads/2020/06/HCCA-Co-design-position-statement-Final.pdf [accessed 2022-04-04]
93. Zhang C. Who bypasses the Great Firewall in China? First Monday 2020 Mar 22;25(4):10256 [FREE Full text] [doi: 10.5210/fm.v25i4.10256]

Abbreviations

F-K: Flesch-Kincaid test
HONcode: Health on the Net Foundation Code of Conduct
HRWEF: Health-Related Website Evaluation Form
ICC: intraclass correlation coefficient
IRR: interrater reliability
QCSS: Quality Component Scoring System
REDCap: Research Electronic Data Capture
SAM: Suitability Assessment of Materials
SMOG: Simple Measure of Gobbledygook

Edited by R Kukafka; submitted 11.04.22; peer-reviewed by Q Zhou, B Nievas Soriano; comments to author 29.05.22; revised version received 03.06.22; accepted 15.07.22; published 07.10.22.

Please cite as:
Jawad D, Cheng H, Wen LM, Rissel C, Baur L, Mihrshahi S, Taki S
Interactivity, Quality, and Content of Websites Promoting Health Behaviors During Infancy: 6-Year Update of the Systematic Assessment
J Med Internet Res 2022;24(10):e38641
URL: https://www.jmir.org/2022/10/e38641
doi:10.2196/38641
PMID:36206031

Original Paper

Quality Assessment of Hypertension Treatment–Related Information on WeChat: Cross-sectional Study

Yuting Yang[1], BS; Mengchi Hou[2], MD; Xue Gong[1], BS; Rui Guo[1*], PhD; Xing Lin Feng[3*], PhD; Rui Tian[4*], MLS

[1]School of Public Health, Capital Medical University, Beijing, China

[2]China Aerospace Science & Industry Corporation 731 Hospital, Beijing, China

[3]Department of Health Policy and Management, School of Public Health, Peking University, Beijing, China

[4]Capital Medical University Library, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Rui Guo, PhD
School of Public Health
Capital Medical University
No 10 Youanmenwai, Xitoutiao
Beijing, 100069
China
Phone: 86 01083911573
Email: guorui@ccmu.edu.cn

Abstract

Background: The WeChat platform has become a primary source for medical information in China. However, no study has been conducted to explore the quality of information on WeChat for the treatment of hypertension, the leading chronic condition.

Objective: This study aimed to explore the quality of information in articles on WeChat that are related to hypertension treatment from the aspects of credibility, concreteness, accuracy, and completeness.

Methods: We searched for all information related to hypertension treatment on WeChat based on several inclusion and exclusion criteria. We used 2 tools to evaluate information quality, and 2 independent reviewers performed the assessment with the 2 tools separately. First, we adopted the DISCERN instrument to assess the credibility and concreteness of the treatment information, with the outcomes classified into five grades: *excellent*, *good*, *fair*, *poor*, and *very poor*. Second, we applied the Chinese Guidelines for Prevention and Treatment of Hypertension (2018 edition) to evaluate the accuracy and completeness of the article information with regard to specific medical content. Third, we combined the results from the 2 assessments to arrive at the overall quality of the articles and explored the differences between, and associations of, the 2 independent assessments.

Results: Of the 223 articles that were retrieved, 130 (58.3%) full texts were included. Of these 130 articles, 81 (62.3%) described therapeutic measures for hypertension. The assessment based on the DISCERN instrument reported a mean score of 31.22 (SD 8.46). There were no articles rated *excellent* (mean score >63); most (111/130, 85.4%) of the articles did not refer to the consequences—in particular, quality of life—of no treatment. For specific medical content, adherence to the Chinese Guidelines for Prevention and Treatment of Hypertension was generally low in terms of accuracy and completeness, and there was much erroneous information. The overall mean quality score was 10.18 (SD 2.22) for the 130 articles, and the scores differed significantly across the 3 types (P=.03) and 5 sources (P=.02). Articles with references achieved higher scores for quality than those reporting none (P<.001). The results from the DISCERN assessment and the medical content scores were highly correlated (ρ=0.58; P<.001).

Conclusions: The quality of hypertension treatment–related information on the WeChat platform is low. Future work is warranted to regulate information sources and strengthen references. For the treatment of hypertension, crucial information on the consequences of no treatment is urgently needed.

(J Med Internet Res 2022;24(10):e38567) doi:10.2196/38567

KEYWORDS

quality assessment; hypertension; WeChat; DISCERN instrument

Introduction

Background

Hypertension is a public health challenge because of its high risk for cardiovascular disease, which is the top cause of morbidity and mortality worldwide [1,2]. In 2019, an estimated 1.28 billion adults aged 30 to 79 years worldwide—approximately 32% to 34% of the global population—were diagnosed with hypertension [3]. In China, nearly half of the adults aged 35 to 75 years were diagnosed with hypertension as of 2017, with medication adherence and control rates <50% and <20%, respectively [4]. Hypertension is a chronic condition that needs lifelong treatment, and the treatment includes two aspects: health management (condition monitoring, lifestyle intervention, control of complications, etc) and taking medications to control rising blood pressure [5,6]. It has been revealed that awareness is the first step for devising appropriate management [7], with the detection rate a key factor that affects treatment and control [8,9]. For patients who are aware of their condition, treatment-related information, including general illness information and treatment choices, is a major concern. Patients also want to play a more active role in decision-making to ease anxiety [10]. Nevertheless, they usually encounter difficulties in finding relevant and easy-to-understand information.

The internet is the first source of medical information for the public as well as patients because of its speed and cost-effectiveness [11]. Toward the end of 2021, China had 1.03 billion internet users, the largest population of *netizens* in the world [12]. WeChat is the primary social media platform for Chinese netizens, equivalent to Facebook for other international community members and providing similar service models. Social media platforms have a convenient search function. On the basis of related keywords, one can retrieve articles, videos, and almost anything one wants [13]. WeChat was launched in 2011, and by June 2022, the monthly WeChat active users had reached 1.3 billion. With >700,000 articles posted daily [14], WeChat has become the most important information source for the Chinese public. Zhang et al [15] found that 98.35% of the participants reported that they had seen health information via WeChat, and WeChat was one of the most popular choices (63.26%) for obtaining health information in China. Despite the benefits, the health information obtained via WeChat has some limitations, with concern about information quality being the most mentioned [16]. On WeChat, the information sources are numerous and unclear, which has resulted in problems of questionable credibility and inaccuracy [17]. Meanwhile, the health literacy of the general population in China is low [10], because of which low-quality health information can lead to harmful behavior. Therefore, it is critical to evaluate the quality of hypertension-related information on WeChat. We found that only 1 study had been conducted to assess the quality of hypertension-related information provided on traditional websites [18]. However, no studies are available assessing the quality of hypertension-related information on WeChat.

Objectives

DISCERN is the most widely used instrument for assessing health-related information and videos, and it is particularly relevant to health-related topics and web-based resources for patient education [19]. Literature is emerging that combines the results from the DISCERN tool and other ratings of web-based references on specific professional content based on clinical guidelines [18,20,21]. This study aimed to assess the quality of information in hypertension treatment–related articles oriented to the general population on WeChat. We adopted the DISCERN instrument to assess the credibility and concreteness of the treatment information and then applied the Chinese Guidelines for Prevention and Treatment of Hypertension (2018 edition; hereinafter referred to as the Hypertension Guidelines) [22] to evaluate the accuracy and completeness of the specific medical content of the treatment information. We combined the results from the 2 sources to report on the overall quality of the articles to comprehensively evaluate the quality of information on WeChat. We believe that this is the first report on the quality of information in hypertension treatment–related articles on WeChat.

Methods

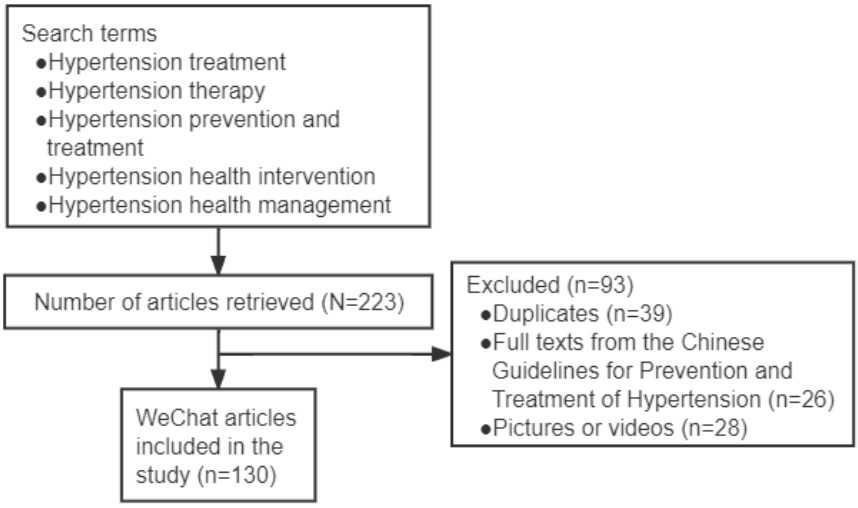
Search Strategy and Data Extraction

In this study, we entered the terms “高血压治疗 (hypertension treatment),” “高血压疗法 (hypertension therapy),” “高血压防治 (hypertension prevention and treatment),” “高血压健康干预 (hypertension health intervention),” and “高血压健康管理 (hypertension health management)” into WeChat for retrieval of relevant articles. To ensure that the articles included in the study matched the research aims, we used certain inclusion and exclusion criteria. The inclusion criteria were as follows: (1) articles focusing on information related to hypertension treatment and health management and (2) articles covering hypertension treatment and therapy. The exclusion criteria were as follows: (1) duplicated articles, (2) articles providing full texts from the Hypertension Guidelines, and (3) articles presented only in picture or video format.

We included relevant data in this study up to the date of information collection, namely May 30, 2021. In Figure 1, we provide some examples of the retrieval strategy and results. After the screening, we included 130 articles for further data extraction and analysis. Figure 2 illustrates the search and screening flow for the articles. We extracted the essential information for each article and its source, including the article title, publication date, numbers of views and *likes*, type of treatment mentioned, uploader (governmental organization vs individual, etc), and references. The extracted data were recorded in Excel (Microsoft Corp).

Figure 1. Examples of the search for articles in WeChat public accounts. (A) Retrieval strategy for hypertension treatment related to the keyword "高血压治疗 (hypertension treatment)" in the WeChat app. (B) Retrieval results. (C) Information provided in the WeChat public accounts. Retrieval date: May 30, 2021.

Figure 2. Search and screening flow for hypertension treatment–related articles.



Ethical Considerations

Institutional review board approval was not required for this study since all information was freely available online. The "articles" were defined as being any piece of open access published writing, excluding personal blogs, editorials, and commentaries.

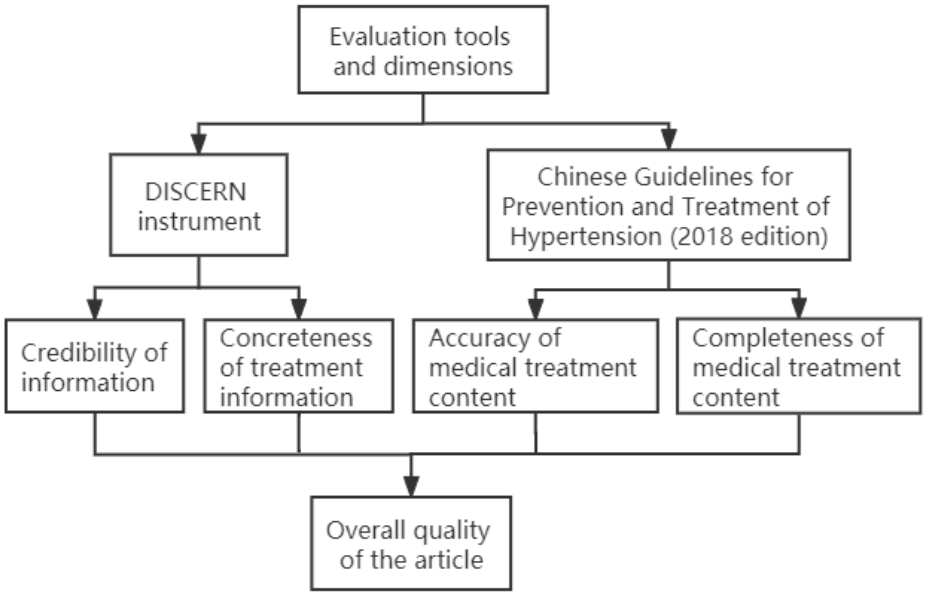
Evaluated Dimensions and Methods

Overview

We measured two aspects of hypertension treatment–related articles on WeChat: the quality of information and the content, to evaluate which we used 2 metrics. First, we adopted the DISCERN instrument, which assesses the credibility and concreteness of written consumer health information with regard to treatment choices. Second, we applied the Hypertension Guidelines as a supplement to evaluate the accuracy and completeness of the specific medical content in the article, which

provided a more granular assessment of the quality of information as it pertains to hypertension treatment. We provide details of the tools in the sections that follow. Third, we combined the results from the 2 assessments to arrive at the overall quality of the articles and explored the differences between, and associations of, the 2 independent assessments (Figure 3). Two researchers assessed each article and scored both instruments. The intraclass correlation coefficient (ICC) was used to measure consistency [23]. The ICC value is between 0 and 1. It is generally acknowledged that ICC>0.80 indicates strong consistency, ICC from 0.80 to 0.41 indicates medium consistency, ICC<0.40 indicates poor consistency, and ICC<0 is considered no consistency [24,25].

Figure 3. Evaluation tools and dimensions.



DISCERN: Assessment of Credibility and Concreteness of Treatment Information

To rate the quality of the information, we adopted the DISCERN instrument. The DISCERN handbook indicates that both professionals and the general population can use the instrument, and a study has confirmed that professionals judge health information similarly to laypersons using DISCERN [26]. The handbook is available on the DISCERN website.

The instrument consists of 16 questions divided into 3 parts (the overall score ranges from 16 to 80) [26]. Part 1 (questions 1 to 8) assesses the credibility of the information, part 2 (questions 9 to 15) focuses on the concreteness of treatment information, and part 3 (question 16) is an overall quality rating [27]. In this study, an article could score up to 80 points on all 16 questions, with up to 40 points for the questions addressing the credibility of information (questions 1-8), up to 35 points for the questions addressing treatment choices (questions 9-15), and up to 5 points for question 16. High scores indicate high quality. For describing and distinguishing the DISCERN scores significantly, we adopted the approach used in a previous study and categorized scores of 63 to 80 as *excellent*, 51 to 62 as *good*, 39 to 50 as *fair*, 27 to 38 as *poor*, and 16 to 26 as *very poor* [18,28]. YY and MH performed the scoring and used the ICC to measure consistency.

Hypertension Guidelines: Assessment of Accuracy and Completeness of Medical Treatment Content

The DISCERN tool can be used for any health-related content area and, thus, is not specific to hypertension [29]. Therefore, we used the Hypertension Guidelines as a supplement to evaluate the accuracy and completeness of the specific medical content in the included articles. We referred to the DISCERN scoring criteria and developed the content evaluation criteria to maintain consistency and comparability with the DISCERN tool [30]. With regard to accuracy, we chose the following categories: completely accurate (5 points), partially accurate (3-4 points), not very accurate but containing no errors in the information (2 points), and wildly inaccurate and containing misinformation (1 point) [25]. For completeness, the Hypertension Guidelines mentions 6 aspects of hypertension treatment [22]: (1) hypertension treatment goals, (2) lifestyle intervention, (3) medical treatment, (4) instrument intervention, (5) management of related risk factors, and (6) treatment of hypertension in special populations. On the basis of the coverage of these 6 key points, we developed the following categories: all 6 key points mentioned (5 points), 4 to 5 key points (4 points), 3 key points (3 points), 1 to 2 key points (2 points), and no mention of any of the key points (1 point). YY and MH performed the scoring and used the ICC to measure consistency.

Overall Article Quality

In general, we combined the DISCERN tool and the Hypertension Guidelines to measure the overall quality of the article. First, we calculated the mean scores of part 1 and part 2 of the DISCERN tool with regard to the credibility and concreteness of information about treatment choices, respectively. Second, we used the 2-part mean scores of accuracy and completeness for the medical treatment content evaluation, reflecting the quality specifically for hypertension. Then, we added the 4-part scores to arrive at the overall quality score for exploration of the quality differences (Figure 3).

Exploring the Quality Differences

Considering the different value propositions, for the comparison, we identified 3 categories of articles, 5 types of sources, and 2 kinds of articles according to whether there were references. First, we combined the treatment aspects of the Hypertension Guidelines and divided the articles into 3 categories: (1) therapeutic measures, (2) lifestyle intervention, and (3) scientific or frontier knowledge (introduction of new drugs, etc). Second, we classified the articles' uploaders into five main categories: (1) governmental organizations, (2) commercial organizations, (3) medical institutions, (4) news or media organizations, and (5) individuals. Third, we divided the articles into 2 kinds according to whether there were references. The purpose was to explore the differences between them in terms of article quality. Detailed information of each uploader was shown in its public account, including the name, time of upload, and institution type (Figure 1).

Statistical Analysis

We used Excel 2019 (Microsoft Corp) for data collection and SPSS software (version 26.0; IBM Corp) for analysis. Data were presented as frequencies and percentages or means and SDs as appropriate. Regarding the evaluation scores, we used the ICC to ascertain the interrater agreement with regard to the exploration of quality differences. Kruskal-Wallis tests were used to determine statistically significant differences between 2 groups or among >2 groups of independent variables. The correlations among the DISCERN scores, content scores, number of views, and number of *likes* were evaluated using Spearman correlation analysis. $P<.05$ was considered statistically significant.

Results

Characteristics of the Articles

In this study, the search retrieved 223 articles, of which we included 130 (58.3%) for analysis according to the inclusion and exclusion criteria (Figure 2). In terms of the treatment information types, 62.3% (81/130) of the articles related to therapeutic measures, 26.9% (35/130) referred to lifestyle intervention, and 10.8% (14/130) involved scientific or frontier knowledge (Table 1). With regard to the uploading source, the majority of the articles had been posted by commercial organizations (78/130, 60%), followed by individuals (29/130, 22.3%), medical institutions (10/130, 7.7%), news or media organizations (9/130, 6.9%), and governmental organizations (4/130, 3.1%; Table 1). Only 13.1% (17/130) of the articles provided references. In addition, 89.2% (116/130) of the articles adopted various marketing strategies for promotion of the content. In Figure 4, we provide examples of these different marketing strategies.

With regard to the number of views, the median was 2929 (range 7 to >100,000); the minor articles (1/130, 0.8%) had been read only 7 times, and only 0.8% (1/130) of the articles had been read >100,000 times. In terms of the *likes* received, the median was 9.5 (range 0-951); >1 article had not received a single *like*, whereas the highest number of *likes* for an article was 951 (Table 1).

Table 1. Characteristics of articles related to hypertension treatment on WeChat (N=130).

Variable	Values
Category, n (%)	
Information category	
Therapeutic measures	81 (62.3)
Lifestyle intervention	35 (26.9)
Scientific or frontier knowledge	14 (10.8)
Uploading source	
Commercial organizations	78 (60)
Individuals	29 (22.3)
Medical institutions	10 (7.7)
News or media organizations	9 (6.9)
Governmental organizations	4 (3.1)
Reference source	
Yes	17 (13.1)
No	113 (86.9)
Metrics, median (range)	
Number of views	2929 (7-100,000)
Number of *likes*	9.5 (0-951)

Figure 4. Examples of different marketing strategies appended to the end of articles.

Evaluated Results

DISCERN: Information Credibility and Concreteness of Treatment Information

The complete table with the average scores of the 130 articles derived by using the 16 questions of the DISCERN instrument is available in Multimedia Appendix 1. Overall, the quality was poor based on the credibility and concreteness of treatment information. According to the DISCERN scale, no article was *excellent* in terms of the information provided; 3.1% (4/130), 18.5% (24/130), and 44.6% (58/130) of the articles were *good*, *fair*, and *poor*, respectively. Besides, 33.8% (44/130) of the articles obtained an abysmal score. An article could score up to 80 points on 16 questions, but the mean score of these 130 WeChat articles was 31.22 (SD 8.46; median 30.00; range 16-58). The ICC was between 0.69 and 0.97, indicating an acceptable consistency.

When part 1 (questions 1-8) and part 2 (questions 9-15) scores are compared, the former comes off slightly better than the latter. In part 1, the article's information credibility was evaluated and received a mean score of 16.58 (SD 4.86). "Are the aims clear?" (question 1) received the highest mean score: 2.87 (SD 0.76). "Does it refer to areas of uncertainty?" (question 8) performed the worst with a mean score of 1.38 (SD 0.79). Part 2 assessed the concreteness of treatment information; the mean score was 12.25 (SD 3.85), indicating that the information provided on treatment choices was generally poor, and this was particularly notable in "Does it describe how the treatment choices affect the overall quality of life?" (question 13), which received a mean score of 1.18 (SD 0.46). Multimedia Appendix 1 shows the detailed scores.

Hypertension Guidelines: Accuracy and Completeness of Medical Treatment Content

With regard to the accuracy and completeness of the hypertension treatment–related medical content, the former performed better than the latter. The ICCs were 0.77 and 0.86, respectively, indicating an acceptable consistency. With regard to article accuracy, the average score was 3.43 (SD 0.79); we found that some (3/130, 2.3%) of the articles contained typographical errors, erroneous information, or areas of content that needed improvement. For completeness, the average score was 2.94 (SD 0.81); most (98/130, 75.4%) of the articles lacked key points on hypertension treatment content according to the Hypertension Guidelines, resulting in the incompleteness of content. In Table 2, we provide some typical examples of content deficiencies or scientific content inaccuracies and suggestions for improvement. Table 3 shows the detailed scores.

Table 2. Typical examples of assessment of the content of some articles on hypertension.

Article title	Areas of deficiencies	Suggestions for improvement
Hypertension prevention and treatment: traditional Chinese therapy has a good remedy, why not try it?	• The title does not match the content. The title emphasized traditional Chinese therapy, whereas the content mainly focused on Western medicine. • There was no significant description of how each therapy works, and there was a lack of concrete critical information, reducing the effect and meaningfulness of the article.	• As the title of the article emphasizes the benefit of using traditional Chinese therapy to treat hypertension, the content should match the title, and the article should describe traditional Chinese therapy in detail. • The article should provide detailed descriptions of drugs and therapies, clarifying the effects of using traditional Chinese therapy and indications for its use; for example, how it works, the benefit, the risk of each treatment, and what would happen if no treatment is used.
How is the hypertension treatment plan developed?	• Hypertension treatment is primarily based on the Chinese Guidelines for Prevention and Treatment of Hypertension, but concreteness was lacking. For hypertension medical treatment, the article only provided a brief statement. • The article did not indicate the source of information.	• Copying and presenting the Chinese Guidelines for Prevention and Treatment of Hypertension is a good idea, but it is better to extract concrete information related to the main idea of the article, keeping in mind the completeness of key content. • Indicating the source of information is good practice and so is providing a link to the reference in the article, as is done in research papers.
Treatment of hypertension (posted by an individual, Xuejie Han)	• The article uses colloquial language, it is verbose, and it even contains wrongly written characters. • With regard to the source, the article indicates that "This text is from the network," but the lack of specifics only raises questions about the scientific and safety issues of the content.	• The writer should avoid using colloquial language, keep sentences simple, and be serious about avoiding incorrect and inaccurate words. • It is better to reference authoritative books, papers, or resources.

Table 3. The mean scores of the articles as evaluated using the DISCERN instrument and the Chinese Guidelines for Prevention and Treatment of Hypertension.

Content evaluated	Scores, mean (SD)
Part 1 (credibility of information)	2.07 (0.61)
Part 2 (concreteness of treatment information)	1.75 (0.55)
Accuracy of treatment information[a]	3.43 (0.79)
Completeness of treatment information[b]	2.94 (0.81)
Overall quality	10.18 (2.22)

[a]Intraclass correlation coefficient=0.77.

[b]Intraclass correlation coefficient=0.86.

Overall Article Quality

First, in terms of the mean DISCERN scores, part 1 (credibility of information) and part 2 (concreteness of treatment information) scored 2.07 (SD 0.61) and 1.75 (SD 0.55), respectively. Second, in terms of the mean Hypertension Guidelines scores, the accuracy and completeness of the medical treatment content scored 3.43 (SD 0.79) and 2.94 (SD 0.81), respectively. Third, the mean score of the overall quality of the articles was 10.18 (SD 2.22; Table 3).

Comparison of Treatment Information Types, Uploading Sources, and Availability of References

We identified 3 categories of articles and 5 types of sources and divided the articles into 2 kinds according to whether they provided references. We chose the overall quality, DISCERN, and medical content scores for comparison. First, there were significant differences among the 3 types (P=.03), primarily because of the differences in medical treatment content quality (P=.02). Second, statistically significant differences could be observed in the overall quality among the 5 sources (P=.02), mainly because of the differences in DISCERN-evaluated quality (P=.02). By contrast, there were no statistically significant differences in the medical content quality scores (P=.10). Governmental institutions scored the highest (mean 11.01, SD 1.36), and individuals scored the lowest (mean 9.08, SD 2.13). Table 4 shows the results. Third, we compared the articles' quality in terms of whether they provided references, and the results showed statistically significant differences (P<.001) between articles that provided references and those that did not. The mean score of articles that provided references was significantly higher than those that did not, 12.90 (SD 1.83) and 9.78 (SD 1.81), respectively.

Table 4. Comparison of the DISCERN scores of information categories, uploading sources, and reference sources.

Item	Score, mean (SD)[a]	*P* value[b]	*P* value[c]	*P* value[d]
Information category		.03	.42	.02
Therapeutic measures	10.51 (2.44)			
Scientific or frontier knowledge	9.91 (2.14)			
Lifestyle intervention	9.54 (1.58)			
Uploading source		.02	.02	.10
Governmental institutions	11.01 (1.36)			
News or media organizations	10.87 (1.34)			
Commercial organizations	10.47 (2.35)			
Medical institutions	10.19 (1.50)			
Individuals	9.08 (2.13)			
Reference source		<.001	<.001	<.001
Yes	12.90 (1.83)			
No	9.78 (1.81)			

[a]The Kruskal-Wallis test was used as a conservative test when determining significance for continuous variables, given the nonnormality of some data. The mean (SD) values refer to the overall quality score.

[b]*P* value is applicable to the overall quality score.

[c]*P* value is applicable to the mean score of DISCERN part 1 and part 2 assessments.

[d]*P* value is applicable to the mean score of accuracy and completeness of medical treatment content.

Quality Assessment and Correlation With Numbers of Views and Likes

Significant correlations were observed between the DISCERN score and the medical content score (*P*<.001). This demonstrated that if the credibility and concreteness of treatment information were excellent, the medical accuracy, completeness, and overall quality were likely to be better; in other words, the DISCERN tool and the Hypertension Guidelines can validate each other. Meanwhile, there was a significant correlation among credibility and accuracy, the concreteness of treatment information, and completeness, which revealed that the evaluations performed using the corresponding parts of the 2 tools were consistent. By contrast, there was no significant correlation among the DISCERN score, the number of views (*P*=.63), and the number of *likes* (*P*=.23); the content score results were similar (*P*=.10 and *P*=.11), which means a good-quality article does not necessarily receive a high number of views (Table 5).

Table 5. *P* values of correlation of DISCERN score, content score, and numbers of views and likes.[a]

	DISCERN score	Content score	Number of views	Number of *likes*
DISCERN score	<.001[b]	—[c]	—	—
Content score	<.001[d]	<.001[b]	—	—
Number of views	.63	.10	<.001[b]	—
Number of *likes*	.23	.11	<.001[e]	<.001[b]

[a]Credibility and accuracy: *P*<.001; concreteness of treatment information and completeness: *P*<.001.

[b]Spearman correlation coefficient=1.00

[c]Not applicable.

[d]Spearman correlation coefficient=0.58.

[e]Spearman correlation coefficient=0.63.

Discussion

Principal Findings

This study has provided the first report on the quality of information in hypertension treatment–related articles on WeChat. The evaluation outcomes from the two sources, that is, the DISCERN and the Hypertension Guidelines, show high correlations and suggest valid results. The overall quality of hypertension treatment–related information on WeChat was poor in terms of credibility, concreteness, accuracy, and completeness. Quality scores differed significantly among the 3 types of articles and 5 information sources, revealing the significance of different value propositions. Articles reporting references were of better quality than those that did not provide

references. Our findings and methods have important implications in an era when people increasingly use social media to obtain health-related information.

Comparison With Prior Work

Prior studies have investigated the quality of web-based information with regard to different diseases and platforms; for example, the study by Azer et al [31] evaluated the quality of information on the internet about inflammatory bowel disease (mean 42.2, SD 10.7), and the study by Kaicker et al [27] evaluated the quality of information on chronic pain (mean 55.9, SD 13.6). We only discovered 1 study evaluating hypertension-related information quality on websites, with a DISCERN score of 45.94, and only 1 of these websites was excellent [19]. The score was higher than ours, probably because WeChat's articles were more subjective and had been uploaded by random sources [31]. With regard to social media platforms, researchers have investigated the quality of YouTube videos about eczema treatment (mean 30.6) and meningioma treatment (mean 36.4, SD 14.0) [32,33], as well as the quality of treatment of rare diseases on WeChat (mean 30.27, SD 7.20). These scores are similar to that obtained in this study (mean 31.22, SD 8.46). Overall, websites performed better than social media platforms. This might be because user barriers as well as barriers to publishing on social media platforms are low because these platforms encourage everyone to participate and share content [34]. However, we must consider the particularities of medical health information, which differs from other types of information [35]. Therefore, it is essential to encourage medical professionals, scientific researchers, and those who have received professional certification to provide health-related content and articles. These people should do more to popularize science and meet the general population's needs.

Articles of Good or Excellent Quality Are Rare

Generally, the quality of hypertension treatment–related information on WeChat was found to be poor in terms of credibility, concreteness, accuracy, and completeness, which is not helpful for the general population. In the DISCERN evaluation, no article was found to be *excellent*, and only 3.1% (4/130) of the articles were rated *good*. This finding is consistent with prior studies on health information on websites and YouTube; for example, the study by San Giorgi et al [36] found that only 2% of the websites evaluated provided *good* content, and the study by Śledzińska et al [33] found that only 4.9% of the YouTube videos assessed were rated *good*. Our study revealed that the credibility of information generally scored higher than the concreteness of treatment information. In other words, compared with the article's credibility, the concreteness of treatment information is harder to achieve. With regard to credibility, "Are the aims clear?" (question 1) received the highest mean score, which was probably related to the emphasis on patient-centered health information services and the need for clear goals in the process of information dissemination [37,38]. This revealed that the match of title and content of hypertension treatment–related information is relatively acceptable. "Does it refer to areas of uncertainty?" (question 8) performed the worst; although most (100/130, 76.9%) of the articles included descriptions of risks and benefits, they failed to mention the uncertainty regarding treatment information. With regard to the concreteness of treatment information, "Does it describe how the treatment choices affect the overall quality of life?" (question 13) received the lowest score, indicating that most articles did not refer to the consequences—in particular, quality of life—of no treatment. Thus, the performance with regard to question 13 was similar to that with regard to question 8, reflecting the lack of concreteness, resulting in incomplete information.

As for the content scores according to the Hypertension Guidelines, most (98/130, 75.4%) of the articles lacked key points on hypertension treatment and provided only 2 to 3 key points, leading to the incompleteness of medical treatment content. Medication and lifestyle interventions are frequently referred to in the articles, probably because pharmacological and nonpharmacological interventions (health management) are common ways to treat hypertension [39]. By contrast, the instrument intervention was hardly ever mentioned, probably because of insufficient evidence regarding the efficacy and safety of this method [22].

Nowadays, the supervision of articles published on WeChat mainly concerns legalities, such as network security and legality of the content. Information quality is not yet a concern, and specific measures to ensure information quality are lacking [40,41]. For this, the Health On the Net code of conduct for medical and health websites (HONcode) can provide some references for improvement measures. The HONcode stipulates that all medical advice must come from medical professionals to ensure the authority and accuracy of the information [42]. For the WeChat platform, we need to consider the professional nature of medical health information. The government must strictly review the author's qualifications as well as the content before publication. In addition, we found that 89.2% (116/130) of the articles had adopted various marketing strategies for promotion of the content, and previous studies have also pointed out this problem. This trend of commercial advertisements disguised as supposedly harmless referral links can become an issue [43]; for example, some publishers exaggerate illness symptoms and product functions to persuade more people to buy drugs and commodities. It is easy to persuade an unsuspecting public that these drugs and products are good for them, but the consequences can be serious in terms of health risks. Therefore, the government must enact strict laws against false advertising with regard to web-based medical information and recommend credible information sources to the public [44].

Governmental Sources Provided High-Quality Information but Were Lacking in Motivation

In this study, significant differences could be observed among the 3 types of articles and 5 uploading sources in terms of overall quality. With regard to the article types, the differences in quality were mainly due to the quality of the medical treatment content. The adherence to the Hypertension Guidelines was low, the articles lacked key treatment management points, and provided incomplete information. Importantly, with regard to the uploading sources, we found that governmental sources scored significantly higher than individuals. This finding was consistent with prior studies, which suggests that governmental

institutions are more likely to publish high-quality information [28]. Presumably because the teams from governmental institutions are highly specialized and knowledgeable, they are more cautious and responsible about what they publish [45]. However, governmental institutions only uploaded 3.1% (4/130) of the articles, which is an indication of poor motivation. As governmental institutions uploaded an insufficient number of articles, we could not arrive at a conclusion regarding their overall performance in the quality score. We know that science is not supposed to be a popularity contest, but governments should exert more effort to disseminate accurate and complete information via social media to ameliorate the negative health consequences of misinformation [46]. By contrast, individuals accounted for 22.3% (29/130) of the articles—presenting high motivation—but the overall quality was not excellent. Noticeably, health care promotion demands high professionalism and strictness, and inaccurate content will mislead the public. Allowing general users to publish content related to highly professional subjects might be inappropriate.

Articles With References Were of Higher Quality

Citation resources or references can reflect an article's objectivity to a certain extent [47]; if they are absent, people's judgment regarding the accuracy of the content as well as their understanding of the health information can be directly affected [36]. In this study, we found that the mean score of articles with references was significantly higher than that of those without, and these scores differed significantly. Of concern, it is not common practice to list references in WeChat articles; only 13.1% (17/130) of the articles listed references. Most (92/130, 70.8%) of the articles provided the author's name; a few (20/130, 15.4%) provided the author's name and work credentials. However, an article's quality and credibility cannot be judged only on the basis of this simple information. Prior studies have suggested that the proportion of content with references was low at 10.2% [45]. Lacking citation resources or references was one of the important factors that led to a severe gap between health information and scientific evidence. The HONcode stipulates traceability, meaning that the content should identify the source of information to which readers could refer [42]. If the WeChat platform identifies the source of the information it publishes, the credibility of this information might be improved. Therefore, the government needs to regulate such articles to ensure that they provide references, perhaps by providing links of citation resources at the end of each article. We found that some WeChat public accounts such as *The Lancet* and *Dingxiang Yisheng* had already done this. Furthermore, we discovered that some articles had provided the number of words in the article and expected reading time, which is a good practice.

Correlations Among the DISCERN Score, Content Score, Number of Views, and Number of *Likes*

A valuable finding of this study was that there were significant correlations between the DISCERN score and the content score. This indicates that the hypertension treatment articles were more likely to be accurate and complete if they included information about the benefits and risk factors of treatment, the consequences of no treatment, and reference resources, presumably because providing reliable and complete information necessitates using more words and furnishing a detailed explanation. By contrast, there were no significant correlations among the DISCERN score, number of views, and number of *likes*, similar to previous study findings [34]. We found that the more popular articles (high numbers of views and *likes*) were not associated with high DISCERN scores. We believe that the numbers of views and *likes* could objectively reflect the interest of the audience as well as the effect of the operation of the public account, although this has no relation to the quality of the article [48]. Perhaps some of the articles with a larger audience had marketing value or were emotionally charged, making them more appealing to viewers rather than providing reliable knowledge [33].

Practical Significance

The findings from this study have significant implications for practice. On the one hand, a method that combines the DISCERN tool and Hypertension Guidelines to evaluate information quality is meaningful and comprehensive for social media platforms and future work. On the other hand, if the credibility and concreteness of treatment information were excellent, the medical accuracy, completeness, and overall quality were likely to be better. On the basis of this finding, we advise WeChat users to identify information quality initially according to the following factors: (1) the completeness of the article: if it clearly describes what (the nature of the disease), why (the cause of the disease), and how (treatment) [31]; (2) whether the article provides references or indicates the source of information; and (3) high numbers of views and *likes* do not necessarily mean excellent quality.

Standardizing and managing information from information providers helps to create a harmonious information environment [49]. In this study, we realized the practical significance of the DISCERN tool, and we suggest that authors should consult the DISCERN handbook when writing their article. First, for example, the DISCERN handbook indicates that a good-quality article must have clear aims and achieve the ultimate goal, which is that the title and content should match [27]. In other words, an article should have an appropriate title that expresses the main idea, and the content should be written around this idea. Second, questions 9 to 15 indicate that a good-quality article about disease treatment must focus on the concreteness of the content, such as how each treatment works, the benefits and risks of each treatment, and what would happen if there was no treatment. Third, the DISCERN handbook also reveals that a good-quality article must clarify the source of information; for example, listing the references or sources at the end of the article is a good practice. In brief, if the article is only providing information about a disease, such as hypertension, the author can refer only to DISCERN part 1 (questions 1-8), but if it is also providing information about treatment of the disease, the author should refer to DISCERN part 1 as well as part 2 (questions 1-15). This may be useful with regard to improving the quality of the information.

Limitations and Future Directions

This study has some limitations. First, we only evaluated the quality of information related to hypertension treatment; thus, the range of the included articles was narrow (only 1 topic).

Second, we only focused on 1 platform; there was no comparison among different platforms, leaving an area for further investigation. Future studies should consider including more evaluation dimensions or comparing among different platforms.

Conclusions

This study is the first to analyze WeChat articles oriented to the general population on hypertension treatment, contributing to a better understanding of the available information on hypertension on WeChat. The evaluation outcomes from the two sources, that is, the DISCERN and the Hypertension Guidelines, show high correlations and suggest valid results. The overall information quality of hypertension treatment–related articles on WeChat was poor. Quality scores differ significantly among the 3 types of articles and 5 uploading sources, revealing the significance of different value propositions. Articles that provided references were of better quality than those that did not. Future work is warranted to regulate information sources and strengthen references. For the treatment of hypertension, crucial information on the consequences of no treatment is urgently needed.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (72174131) and the Humanities and Social Science Research Program of the Ministry of Education (21YJA870008).

Authors' Contributions

YY and RG designed the main concepts of this work. YY, MH, and XG performed data collection and wrote this paper. RG, XLF, and RT critically reviewed and revised the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The DISCERN scores for the hypertension treatment information.
[DOCX File , 18 KB - jmir_v24i10e38567_app1.docx]

References

1. He J, Whelton PK. Epidemiology and prevention of hypertension. Medical Clin North America 1997 Sep;81(5):1077-1097. [doi: 10.1016/s0025-7125(05)70568-x]
2. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005;365(9455):217-223. [doi: 10.1016/S0140-6736(05)17741-1] [Medline: 15652604]
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
4. Lu J, Lu Y, Wang X, Li X, Linderman GC, Wu C, et al. Prevalence, awareness, treatment, and control of hypertension in China: data from 1·7 million adults in a population-based screening study (China PEACE Million Persons Project). Lancet 2017 Dec 09;390(10112):2549-2558. [doi: 10.1016/S0140-6736(17)32478-9] [Medline: 29102084]
5. Zhang TJ, Wu SD. 1999 World Health Organization and International Hypertension Association (WHO/ISH) Guidelines for the Treatment of Hypertension [in Chinese]. Advances Cardiovascular Diseases 1999(3):177. [doi: 10.16806/j.cnki.issn.1004-3934.1999.03.019]
6. Yu Z, Chen Y. Interpretation of ISH 2020 International Hypertension Practice Guidelines [in Chinese]. Chinese Rural Med 2022;27(23):24-25. [doi: 10.19542/j.cnki.1006-5180.004626]
7. 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]
8. Feng XL. Undiagnosed and uncontrolled chronic conditions in China: could social health insurance consolidation make a change? Med Care Res Rev 2018 Aug;75(4):479-515. [doi: 10.1177/1077558717690303] [Medline: 29148342]
9. Feng XL, Pang M, Beard J. Health system strengthening and hypertension awareness, treatment and control: data from the China Health and Retirement Longitudinal Study. Bull World Health Organ 2014 Jan 01;92(1):29-41 [FREE Full text] [doi: 10.2471/BLT.13.124495] [Medline: 24391298]
10. Luo A, Xin Z, Yuan Y, Wen T, Xie W, Zhong Z, et al. Multidimensional feature classification of the health information needs of patients with hypertension in an online health community through analysis of 1000 patient question records: observational study. J Med Internet Res 2020 May 29;22(5):e17349 [FREE Full text] [doi: 10.2196/17349] [Medline: 32469318]

11. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
12. CNNIC: The 49th "Statistical Report on Internet Development in China". CNNIC. 2022 Mar 11. URL: http://www.100ec.cn/home/detail--6608634.html [accessed 2022-08-08]
13. Top 10 social software brands in 2022. Maigoo. URL: https://www.maigoo.com/maigoo/1211sjmt_index.html [accessed 2022-08-08]
14. WeChat public account article data report: 90% of original articles read tens of thousands per day. domeet webmaster. 2019 Nov 6. URL: https://page.om.qq.com/page/Or5HrWDZdxmGDlQbHtCxEnaw0 [accessed 2022-08-08]
15. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
16. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
17. Azer SA, Alghofaili MM, Alsultan RM, Alrumaih NS. Accuracy and readability of websites on kidney and bladder cancers. J Cancer Educ 2018 Aug 9;33(4):926-944. [doi: 10.1007/s13187-017-1181-z] [Medline: 28281091]
18. Peng L, Tang G, Yang H, Du Z. Quality evaluation of hypertension health information in network environment. Modern Preventive Med 2017;44(20):3727-3730.
19. Davaris MT, Dowsey MM, Bunzli S, Choong PF. Arthroplasty information on the internet: quality or quantity? Bone Jt Open 2020 Apr;1(4):64-73 [FREE Full text] [doi: 10.1302/2633-1462.14.BJO-2020-0006] [Medline: 33215109]
20. Wu C, Guo RQ, Lv SQ, Sun JJ, Guan ZT, Li Y. Quality evaluation of diabetes health information on Chinese Internet sites [in Chinese]. Chinese J Evidence Based Med 2019;19(10):1250-1254.
21. Wu C, Wu H, Bi JQ, Sun YC. Quality assessment of health information on lung cancer in Chinese websites [in Chinese]. China cancer 2018;27(06):433-437.
22. Chinese Revision Committee of Guidelines for Prevention and Treatment of Hypertension in China. Chinese Guidelines for the Prevention and Treatment of Hypertension (2018 Edition). Chinese J Cardiovascular Med 2019;24(01):24-56.
23. Azer SA, Azer S. Bibliometric analysis of the top-cited gastroenterology and hepatology articles. BMJ Open 2016 Feb 08;6(2):e009889 [FREE Full text] [doi: 10.1136/bmjopen-2015-009889] [Medline: 26857105]
24. Interclass correlation coefficient. zhihu. URL: https://zhuanlan.zhihu.com/p/375131294 [accessed 2021-08-15]
25. Chong YM, Fraval A, Chandrananth J, Plunkett V, Tran P. Assessment of the quality of web-based information on bunions. Foot Ankle Int 2013 Aug 11;34(8):1134-1139. [doi: 10.1177/1071100713481458] [Medline: 23478888]
26. Griffiths KM, Christensen H. Website quality indicators for consumers. J Med Internet Res 2005 Nov 15;7(5):e55 [FREE Full text] [doi: 10.2196/jmir.7.5.e55] [Medline: 16403719]
27. Kaicker J, Borg Debono V, Dang W, Buckley N, Thabane L. Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. BMC Med 2010 Oct 12;8:59 [FREE Full text] [doi: 10.1186/1741-7015-8-59] [Medline: 20939875]
28. Ved R, Cobbold N, Igbagiri K, Willis M, Leach P, Zaben M. Online patient information on Vagus Nerve Stimulation: how reliable is it for facilitating shared decision making? Seizure 2017 Aug;50:125-129 [FREE Full text] [doi: 10.1016/j.seizure.2017.06.009] [Medline: 28651125]
29. Haragan AF, Zuwiala CA, Himes KP. Online information about periviable birth: quality assessment. JMIR Pediatr Parent 2019 Jun 07;2(1):e12524 [FREE Full text] [doi: 10.2196/12524] [Medline: 31518325]
30. Zhang KY, Li H. Research on information quality evaluation index system of health WeChat public platform [in Chinese]. Inform Sci 2017;35(11):143.
31. Azer SA, AlOlayan TI, AlGhamdi MA, AlSanea MA. Inflammatory bowel disease: an evaluation of health information on the internet. World J Gastroenterol 2017 Mar 07;23(9):1676-1696 [FREE Full text] [doi: 10.3748/wjg.v23.i9.1676] [Medline: 28321169]
32. Gorrepati PL, Smith GP. DISCERN scores of YouTube information on eczema treatments. J Am Acad Dermatol 2021 Nov;85(5):1354-1355. [doi: 10.1016/j.jaad.2020.11.007] [Medline: 33181240]
33. Śledzińska P, Bebyn MG, Furtak J. Quality of YouTube videos on meningioma treatment using the DISCERN instrument. World Neurosurg 2021 Sep;153:e179-e186. [doi: 10.1016/j.wneu.2021.06.072] [Medline: 34166826]
34. Baidu homepage. Baidu. URL: https://baike.baidu.com/item/%E7%A4%BE%E4%BA%A4%E5%AA%92%E4%BD%93/1085698?fr=aladdin [accessed 2022-08-08]
35. Particularity of medical service. Blog China. URL: https://net.blogchina.com/blog/article/2239346 [accessed 2022-08-08]
36. San Giorgi MR, de Groot OS, Dikkers FG. Quality and readability assessment of websites related to recurrent respiratory papillomatosis. Laryngoscope 2017 Oct 24;127(10):2293-2297 [FREE Full text] [doi: 10.1002/lary.26521] [Medline: 28233911]

37. Gu D, Yang X, Li X, Jain HK, Liang C. Understanding the role of mobile internet-based health services on patient satisfaction and word-of-mouth. Int J Environ Res Public Health 2018 Sep 10;15(9):1972 [FREE Full text] [doi: 10.3390/ijerph15091972] [Medline: 30201921]
38. Xu XY. The information communication model of WeChat from a perspective on communication. New Media Res 2016;2(21):28-29.
39. Unda Villafuerte F, Llobera Cànaves J, Lorente Montalvo P, Moreno Sancho ML, Oliver Oliver B, Bassante Flores P, et al. Effectiveness of a multifactorial intervention, consisting of self-management of antihypertensive medication, self-measurement of blood pressure, hypocaloric and low sodium diet, and physical exercise, in patients with uncontrolled hypertension taking 2 or more antihypertensive drugs: The MEDICHY study. Medicine (Baltimore) 2020 Apr;99(17):e19769 [FREE Full text] [doi: 10.1097/MD.0000000000019769] [Medline: 32332617]
40. Zhu G, Jin XY, Xu SY. Research on intellectual property protection of WeChat public platform from the perspective of tripartite game [in Chinese]. Yuejiang J 2020;12(3):36.
41. Lv K. Analysis of the government supervision of WeChat public accounts [in Chinese]. Media Forum 2020(5):44.
42. HONcode. Docin. URL: https://www.docin.com/p-373108838.html [accessed 2022-01-01]
43. Mueller SM, Hongler VNS, Jungo P, Cajacob L, Schwegler S, Steveling EH, et al. Fiction, Falsehoods, and Few Facts: Cross-Sectional Study on the Content-Related Quality of Atopic Eczema-Related Videos on YouTube. J Med Internet Res 2020 Apr 24;22(4):e15599 [FREE Full text] [doi: 10.2196/15599] [Medline: 32329744]
44. Li Y, Ye S, Zhou Y, Mao F, Guo H, Lin Y, et al. Web-based medical information searching by Chinese patients with breast cancer and its influence on survival: observational study. J Med Internet Res 2020 Apr 17;22(4):e16768 [FREE Full text] [doi: 10.2196/16768] [Medline: 32301740]
45. Shao C, Guan X, Sun J, Cole M, Liu G. Social media interactions between government and the public: a Chinese case study of government WeChat official accounts on information related to COVID-19. Front Psychol 2022;13:955376 [FREE Full text] [doi: 10.3389/fpsyg.2022.955376] [Medline: 36148095]
46. Moon H, Lee GH. Evaluation of Korean-language COVID-19-related medical information on YouTube: cross-sectional infodemiology study. J Med Internet Res 2020 Aug 12;22(8):e20775 [FREE Full text] [doi: 10.2196/20775] [Medline: 32730221]
47. Fisher JH, O'Connor D, Flexman AM, Shapera S, Ryerson CJ. Accuracy and reliability of internet resources for information on idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2016 Jul 15;194(2):218-225. [doi: 10.1164/rccm.201512-2393oc]
48. Zhang YP. Research on the operation mode of WeChat official accounts of science and technology journals—Based on quantitative analysis of four core science and technology journals [in Chinese]. Chinese J Sci Technol Res 2015;26(5):524-531.
49. Yang F, Shi Q, Yao LY. Evaluation of government social media information quality based on user subject cognition: investigation of government microblog in Chinese. J Inform 2015;34(12):181-185.

Abbreviations

HONcode: Health On the Net code of conduct for medical and health websites
ICC: intraclass correlation coefficient

Edited by G Eysenbach; submitted 08.04.22; peer-reviewed by R Khan, Y Lai; comments to author 18.07.22; revised version received 09.09.22; accepted 23.09.22; published 26.10.22.

Please cite as:
Yang Y, Hou M, Gong X, Guo R, Feng XL, Tian R
Quality Assessment of Hypertension Treatment–Related Information on WeChat: Cross-sectional Study
J Med Internet Res 2022;24(10):e38567
URL: https://www.jmir.org/2022/10/e38567
doi:10.2196/38567
PMID:36287598

Original Paper

Understanding Trust and Changes in Use After a Year With the NHS COVID-19 Contact Tracing App in the United Kingdom: Longitudinal Mixed Methods Study

Cecily Pepper[1*]; Gisela Reyes-Cruz[2*]; Ana Rita Pena[1*]; Liz Dowthwaite[3]; Camilla M Babbage[4]; Hanne Wagner[2,3]; Elena Nichele[3]; Joel E Fischer[2]

[1]Horizon Centre for Doctoral Training, School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[2]Mixed Reality Lab, School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[3]Horizon Digital Economy Research, University of Nottingham, Nottingham, United Kingdom

[4]NIHR MindTech MedTech Co-operative, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Cecily Pepper
Horizon Centre for Doctoral Training
School of Computer Science
University of Nottingham
Triumph Road
Nottingham, NG8 1AW
United Kingdom
Phone: 44 1158232316
Email: cecily.pepper@nottingham.ac.uk

Abstract

Background: Digital contact tracing (DCT) apps have been implemented as a response to the COVID-19 pandemic. Research has focused on understanding acceptance and adoption of these apps, but more work is needed to understand the factors that may contribute to their sustained use. This is key to public health because DCT apps require a high uptake rate to decrease the transmission of the virus within the general population.

Objective: This study aimed to understand changes in the use of the National Health Service Test & Trace (T&T) COVID-19 DCT app and explore how public trust in the app evolved over a 1-year period.

Methods: We conducted a longitudinal mixed methods study consisting of a digital survey in December 2020 followed by another digital survey and interview in November 2021, in which responses from 9 participants were explored in detail. Thematic analysis was used to analyze the interview transcripts. This paper focuses on the thematic analysis to unpack the reasoning behind participants' answers.

Results: In this paper, 5 themes generated through thematic analysis are discussed: flaws in the T&T app, usefulness and functionality affecting trust in the app, low trust in the UK government, varying degrees of trust in other stakeholders, and public consciousness and compliance dropping over time. Mistrust evolved from participants experiencing sociotechnical flaws in the app and led to concerns about the app's usefulness. Similarly, mistrust in the government was linked to perceived poor pandemic handling and the creation and procurement of the app. However, more variability in trust in other stakeholders was highlighted depending on perceived competence and intentions. For example, Big Tech companies (ie, Apple and Google), large hospitality venues, and private contractors were seen as more capable, but participants mistrust their intentions, and small hospitality venues, local councils, and the National Health Service (ie, public health system) were seen as well-intentioned but there is mistrust in their ability to handle pandemic matters. Participants reported complying, or not, with T&T and pandemic guidance to different degrees but, overall, observed a drop in compliance over time.

Conclusions: These findings contribute to the wider implications of changes in DCT app use over time for public health. Findings suggest that trust in the wider T&T app ecosystem could be linked to changes in the use of the app; however, further empirical and theoretical work needs to be done to generalize the results because of the small, homogeneous sample. Initial novelty effects occurred with the app, which lessened over time as public concern and media representation of the pandemic decreased and

normalization occurred. Trust in the sociotechnical capabilities of the app, stakeholders involved, and salience maintenance of the T&T app in conjunction with other measures are needed for sustained use.

(J Med Internet Res 2022;24(10):e40558) doi:10.2196/40558

KEYWORDS

COVID-19; tracing app; digital contact tracing; trust; public health; technology adoption; compliance; longitudinal; mixed methods; thematic analysis; mobile phone

Introduction

Digital Contact Tracing and the COVID-19 Pandemic

Throughout the COVID-19 pandemic, mobile apps enabling digital contact tracing (DCT) became a widespread solution adopted by many countries worldwide for mitigating the spread of SARS-CoV-2 [1,2]. A need for understanding how people feel toward these apps arose once they started being announced and released from 2020 onward. Several studies investigated people's perceptions and attitudes in relation to the acceptance and adoption of DCT [3-6], privacy and ethical concerns [7-9], and the type of messaging used to promote the apps (eg, targeting individualism or collectivism) [7,10]. Although conducted using hypothetical apps and scenarios, early studies found that large percentages of the people consulted said they would use the apps if they became available and that, initially, many people had positive perceptions toward DCT as a medium for minimizing the spread of the virus in society [3,11].

However, after 2 years of the pandemic, the overall uptake of DCT apps remained low, especially because they were mostly voluntary [2]. This imposes a technical challenge given that some studies suggest that the apps need a high uptake rate (from 56% to 95%) to work effectively [12,13]. However, other theoretical and empirical studies found that effects can also be seen at lower adoption levels. For instance, research has estimated that >32% uptake can be helpful in lowering the epidemic to manageable levels when the infection rate is moderate [14]. Other work has estimated that cases could be reduced between 0.8% and 2.3% for every incremental percentage point of app uptake [15]. Nonetheless, regardless of infection rates, higher uptake would represent a higher number of cases averted [15], which makes the case for better understanding DCT app uptake and further incentivizing them.

An intention-behavior gap has been identified in DCT uptake [11], but little is known about the *actual* experiences of people with the apps [16], especially over time, as there have been only few longitudinal studies in this regard [8]. Privacy has been found to be a major concern of people, possibly hindering app uptake [6,17], but the "privacy paradox" has been also acknowledged and investigated, where although people state being highly worried about their privacy, they do not act on their concerns [18]. Thus, more work is needed to understand people's experiences with DCT and other factors, beyond privacy concerns, that may be involved in their adoption and sustained use over time. In the context of DCT, sustained use can be understood in terms of passively engaging with the specific characteristics of the system such as having Bluetooth tracing enabled and taking part in more active aspects of contact tracing such as checking in to public places.

This study builds on previous explorations of trust in DCT apps [19], where trust has been found to be a major factor in their intended adoption, examining in further detail people's experiences and trust perceptions—in relation to a DCT app and its related ecosystem—and their relationship to DCT adoption and use over time. In this study, ecosystem is defined as the set of stakeholders involved in developing, maintaining, and using the Test & Trace (T&T) app. Trust is a complex, context-dependent subject [20], being considered as a mental state felt by people toward others (people, systems, organizations, etc) [21,22]. Regarding technology, trust in stakeholders who control it is often a prerequisite for trust in the technology itself [22]. In the context of DCT, this effect was also shown in research by von Wyl et al [23], where trust in the Swiss government and health authorities was positively correlated with app uptake. Trust therefore plays an important role in technology acceptance uptake.

Technology Acceptance Concepts

There exists a series of models of factors influencing acceptance, with the technology acceptance model (TAM) [24] being one of the most well-known models, which includes factors such as perceived usefulness and perceived ease of use. Specific models for health care informatics applications have also been developed such as the health information TAM, which includes factors such as health status, health beliefs and concerns, and perceived health threat [25]. In this study, we draw from the technology acceptance lifecycle (TAL) by Nadal et al [26], which creates terminologies to take into consideration the temporality of technology acceptance by dividing it into three main stages: before use acceptability (before the first use), initial use acceptance (after the first use but before adoption), and finally, sustained use acceptance (after adoption). Initial use is related to the human-computer interaction literature concept of the novelty effect (NE), which is the set of responses to initially using technology but that does not equate to the long-term use pattern [27], or sustained use under TAL terminology.

Several studies on NE have shown that as it "wears off," many users stop using the technology [28,29]. In an activity tracker long-term use study [29], research showed that only curiosity about the technology and data is not enough to provoke sustained use, which tends to be associated with personal and social motivation as well as gaming motivation. However, studies of the NE in health informatics are related to the user's own health as opposed to individual action for public or collective health, which is the case for DCT in the context of the COVID-19 pandemic, which will be explored in this paper.

Context of the Study

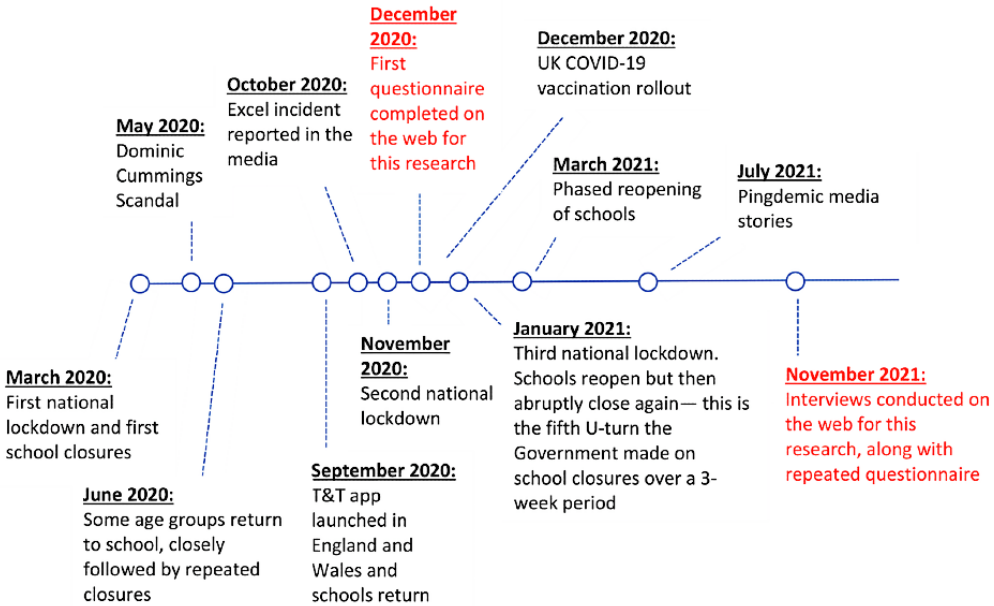
In this study, we focused on the National Health Service (NHS) COVID-19 T&T app, which is the national DCT app for England and Wales. The T&T app was launched in England and Wales on September 24, 2020. The app uses Bluetooth for contact tracing by recording locally on a smartphone the amount of time spent with, and distance between, users. If a user has been in close contact with someone who tests positive, the app will notify the user and give guidance. Furthermore, the T&T app also allows one to check in symptoms, book tests, and input test results as well as checking in to various places such as hospitality venues (eg, pubs or restaurants) by scanning the venue QR code onto the app. In this study, we consider sustained use as engaging with the above features, both in a passive (eg, having Bluetooth tracing on) or more active (eg, checking in at venues) manner. The T&T app had a total of 20.35 million downloads (data from December 2, 2020) at the approximate time of the first survey and 28.76 million (data from November 3, 2021) at the time of the second survey and interview [30].

As the T&T app was designed for use within the United Kingdom, a breakdown of COVID-19–related events, lockdowns, and measures is necessary for context. The first lockdown in the United Kingdom occurred in March 2020, with lockdown measures being legally enforced soon after [31]. In September 2020, the T&T app was launched in England and Wales with the hope to "help control coronavirus (Covid-19) transmission" [32]. Multiple lockdowns followed in November 2020 and January 2021 [31]. Throughout the pandemic, multiple school closures and reopenings occurred, with the UK government making several U-turns in their decisions [33]. The COVID-19 vaccination rollout began in December 2020, starting with the most vulnerable and then gradually extending through age groups and risk levels [34] (see the timeline in Figure 1).

Along with providing context for the lockdowns and measures of the United Kingdom, several media events were also significant in the public's perception of the T&T app. In May 2020, a scandal occurred with the prime minister's chief advisor, Dominic Cummings, who faced public outrage and calls to resign after driving across the country during a UK lockdown [35]. Several months later, in October 2020, shortly after the launch of the T&T app, the media reported an error by Public Health England, who had been using Microsoft Excel to store public health data and an error in formatting resulted in >15,000 unreported cases of COVID-19 [36]. In addition, as the public continued to use the app to check in to venues, a surge in "pings" (notifications that tell the user they need to self-isolate) occurred in July 2021 that was labeled by the media as the "pingdemic." This caused multiple issues with manufacturing and hospitality, especially because there was a legal duty to self-isolate if the user was pinged [37]. The timeline (Figure 1, inspired by the timeline from the Institute for Government [31]) displays the key events as discussed: the 3 national lockdowns, school closures, and reopenings; 3 significant media scandals; and the 2 data collection points for this research.

This longitudinal study builds on the research by Dowthwaite et al [19], which was a quantitative study that investigated attitudes and trust toward the T&T app. To understand this further, this study aims to explore how such attitudes and trust changed or were maintained over time by analyzing research data collected approximately a year apart.

Figure 1. A timeline detailing the key events that are referred to in this research. T&T: Test & Trace; UK: United Kingdom.



Methods

Research Design

This research had a longitudinal design as questionnaire data were collected at 2 time points roughly a year apart, with answers being explored in detail in an interview at the second time point. Therefore, the research had a mixed methods design, gathering quantitative data from the questionnaires and qualitative data from the interviews. The analysis makes use of the quantitative data as a backdrop but emphasizes the qualitative data collected to fulfill the aim of exploring the

reasons for any changes that may have occurred over time in attitudes, use, and trust toward the T&T app.

Ethics Approval

Ethics approval was granted by the research ethics committee of the authors' institutions (approval number: CS-2020-R80 for the interviews and CS-2020-R10 for the questionnaires).

Participants

To study changes in attitudes toward the T&T app, participants who took part in a questionnaire survey in December 2020 were invited to participate in an interview approximately 1 year later to see if and how their question responses may have changed and why. Participants were recruited via email and social media through the authors' personal and professional networks (eg, listserv mailing lists, Twitter, and Facebook) to participate in a web-based questionnaire survey and were asked to consent to being contacted for a follow-up interview. A total of 48 people began the questionnaire, with 40 complete responses. From these respondents, 16 agreed to be contacted. Interviews were conducted with 9 participants who responded when contacted in November 2021. They were aged between 29 and 49 (average age 38, SD 8) years. Of 9 participants, 3 (33%) were men, 4 (44%) were women, and 2 (22%) were nonbinary (Table 1). We did not explicitly ask about occupation, although some participants discussed it in their responses.

Table 1. Basic demographics of participants (N=9).

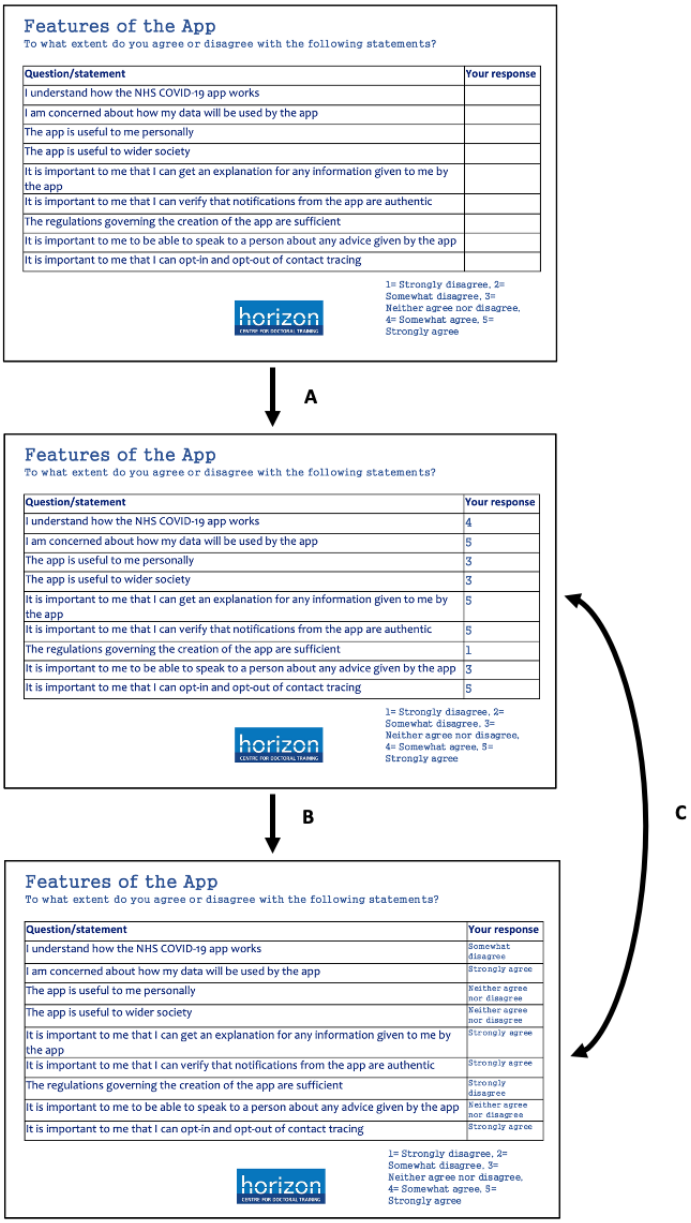
Demographic	Participants, n (%)
Gender	
Man	3 (33)
Woman	4 (44)
Nonbinary	2 (22)
Highest educational level	
Undergraduate degree	1 (11)
Master's degree	7 (78)
Doctorate	1 (11)
Employment status	
Employed full-time	4 (44)
Employed part-time	2 (22)
Student	3 (33)
Ethnicity	
White	8 (89)
Asian	1 (11)
Religion	
None	5 (56)
Christian	2 (22)
Muslim	1 (11)
Other (not specified)	1 (11)

Materials and Procedure

Participants answered the same questionnaire described in an earlier study in this journal [19], at 2 time points, between November 13 and December 23, 2020, when the United Kingdom was between "lockdown 2" and "lockdown 3" and subject to a regional tier system, and between October 25 and November 5, 2021, when most restrictions across the United Kingdom had been lifted. The first questionnaire was administered on the web, whereas the second was administered as part of an interview. At the start of the web-based questionnaire (Multimedia Appendix 1), participants were provided with information and privacy notices and gave informed consent to participate. Questions took the form of either multiple choice or Likert and Likert-like scales, except for a single open-ended question that was included to elicit further comments. The first part of the survey asked participants to indicate what knowledge and experiences they had of COVID-19 and the NHS T&T app; for example, compliance with any requests to self-isolate, whether they had downloaded the app, and if not downloaded, then why not. Those who had downloaded the app were then asked for their reasons for downloading and experiences of using the app. They were then asked about the technology and functionality of the app, including perceived usefulness and ease of use, understanding of how it worked, and the importance of features such as opting in and out of contact tracing. Finally, they were asked about the levels of trust in distinct aspects of the app, including responsibility, security, reliability, functionality, data use, and stakeholders and wider society.

The web-based interview, approximately 11 months later, lasted between 31 and 56 (average time 44) minutes. After confirming their consent to the interview, the interviewer started recording the session and shared their screen with the participants. The questions from the earlier questionnaire were shown to the participants, presented in groups as they were in the original survey, without their previously provided responses (Figure 2). Participants were asked to respond to the questions as they would now, and these responses were added to the slide; then, they were shown their original responses of the year before. Any changes were then probed; for example, why they felt more positive or negative or why they thought they responded a particular way originally but not now. This was done to provide a visual reminder to the participants of the questions and their responses, which could easily be compared between slides, and to provide a focus during the discussion. Following this, they were asked to summarize how their experiences with the T&T system had changed since they filled out the questionnaire, how they feel about the app and T&T, and how their actions had changed over the previous year. They were also asked how their trust in T&T had changed and whether the rollout of vaccines affected this. Finally, they were asked to highlight any specific media stories, events, or other factors that had affected their trust in T&T.

Figure 2. Example of a question block displayed to participants. After being shown the blank slide, they fill in their current answers in November 2021 (A). After this, they are shown their previous responses from December 2020 (B) and they can refer back between slides as visual reminders of their responses during their interviews as indicated by (C). NHS: National Health Service.

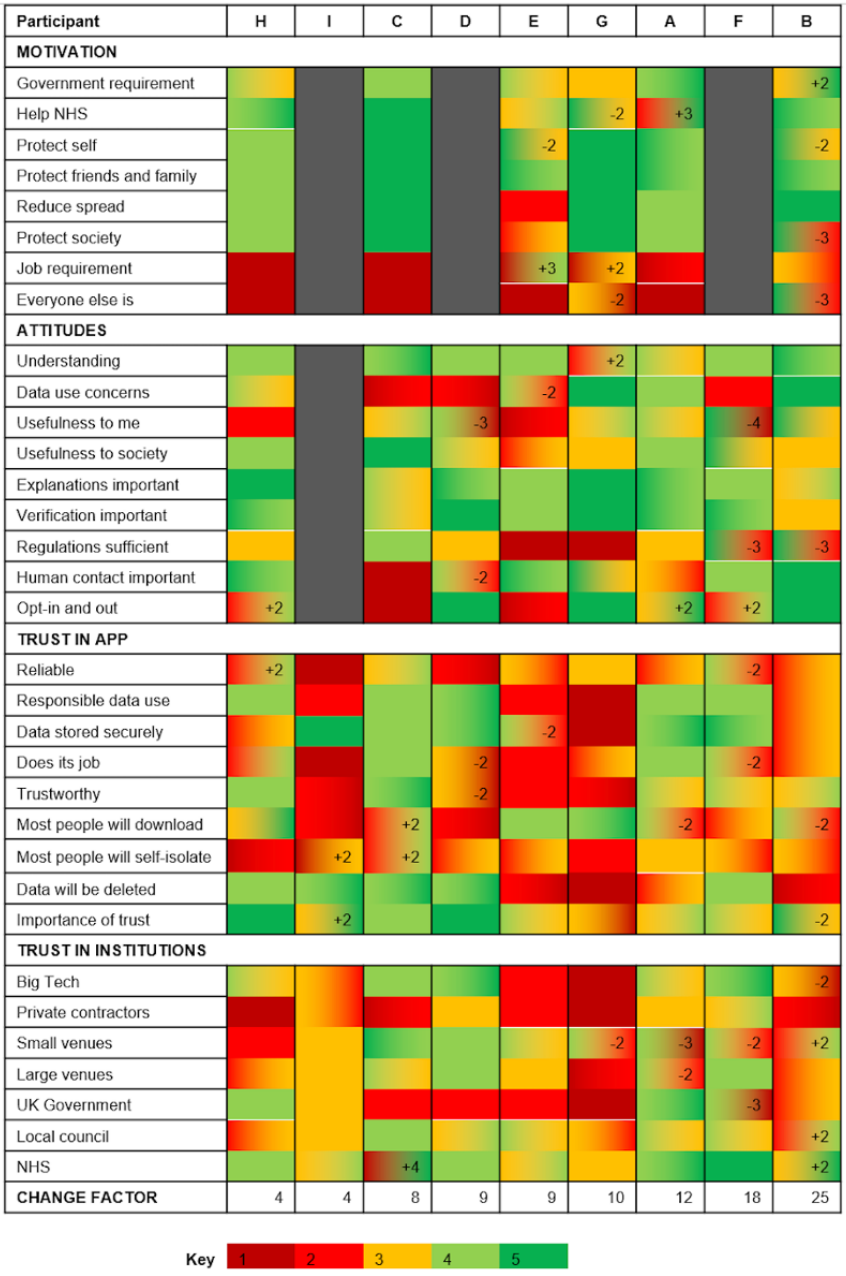


Analysis

Interviews were recorded through Microsoft Teams and transcribed through a 2-step process: first, by generating initial automated transcriptions through Microsoft Stream and then by thoroughly listening to the interview recordings and revising the transcripts to ensure accuracy.

The initial questionnaire data from December 2020 were downloaded to a spreadsheet, and after the interviews, new responses to each question were added. The difference between initial and second responses was calculated numerically for statements where possible (Figure 3). No other summary statistics or any inferential statistics was calculated because we were interested in the within-participant changes in response. Because of the nature of repeated responses, where minor fluctuations in response may be expected day-to-day, it may reasonably be expected that a change of a single point (eg, strongly agree to agree) may not represent an actual change in opinion; only changes of ≥2 points are reported. This means that a participant would have at least changed from strongly agree or disagree to neutral, agree to disagree, or vice versa. A change factor was reported (Figure 3) by calculating the sum of each participant's change points, regardless of the direction, to summarize the amount of general change in trust and perceptions of the T&T app each participant had experienced over time.

Figure 3. Change in questionnaire responses between December 2020 and November 2021 by the participants. Dark red (1) for strongly disagree up to dark green (5) for strongly agree and gray for not applicable (for participants who did not have the app installed at the point of interview). Positive and negative changes of ≥2 points are indicated by integer and color gradient. Participant columns are ordered by ascending change factor, which indicates the sum of ≥2 change points by the participants. NHS: National Health Service; UK: United Kingdom.



The first 3 authors conducted a thematic analysis of the interview data by following the 6 phases established by Braun and Clarke [38]. The data were analyzed from an experiential perspective with the aim to capture and explore people's own perspectives and understanding [39]. The 3 authors familiarized themselves with the data by (1) first correcting 3 interview transcriptions each and then reviewing the whole data set and then (2) inductively generated initial codes aiming to capture

both semantic and latent meanings. The authors used a list of common terms characterizing the interviews for creating the codes, but no predefined concepts were searched for while coding the data. The process was conducted using Microsoft Excel through an index system to track each code generated and match it with the corresponding raw data. The authors then (3) reviewed each other's work once all interview transcriptions were coded and collaboratively searched for candidate themes by collating the codes, which then (4) were reviewed and developed into themes and subthemes. These were (5) named and defined and (6) elaborated at length in writing, selecting relevant data extracts for illustrating them.

The 3 authors kept a reflexive journal throughout and kept ongoing reflective discussions for considering possible biases and personal perspectives in relation to T&T and the overall pandemic experience to be aware of their influence on the analytical process. All 3 authors had different personal experiences of using the T&T app, from never downloading the app, to experiencing flaws, and continued use; hence, the overall perspective on the T&T app felt balanced. The collaborative analysis further enabled a more nuanced understanding of the data, rather than aiming for a consensus about the meaning [38].

Results

Summary of Responses and Themes

The comparison of responses given by the participants on the 2 study time points (December 2020 and November 2021) is summarized as a chart indicating positive, negative, and no changes, as well as the overall change factor (Figure 3). Participants D and F deleted the app between the initial questionnaire study and the interview study; participant I never downloaded the app and did not intend to do so; the remaining participants downloaded the app and still had it on their phones at the time of the interview. Further details about the participants, such as demographics and a summary of individual responses and changes, are provided in Figure 2 and Multimedia Appendix 2.

A total of 6 themes were developed (Multimedia Appendix 3) from the analysis of the 9 interviews. As the interview transcripts were coded inductively, the sixth theme was not entirely relevant to the research aims; therefore, in this study, we concentrated on elaborating 5 themes relevant to the research objectives (as stated in the *Introduction* section). Themes revolved around the T&T app as experienced by participants, who highlighted several *encountered and perceived flaws* (T1) and reported *different degrees of trust* in it (T2). Themes further developed around participants' trust in stakeholders, with reported general low *trust in the government* owing to their poor pandemic handling (T3) and a mix of *trust in other stakeholders* based on their perceived competence and intentions (T4). The final theme revolved around participants' *changing experiences over time* in relation to the T&T app, their general compliance and trust in stakeholders, as well as decreased exposure and reports observed over time (T5). In the following sections, the 5 themes are elaborated on, including participants' quotes and references to their changed or maintained responses (Figure 3).

T1: Flaws of the T&T App Perceived and Experienced by Participants

1a: Lack of Technical Advance and Lack of General Public Uptake

Participants mentioned practical issues related to the proximity measures of the app relying on Bluetooth, including not having it on all the time and its effectiveness in helping track exposure to the virus, especially in complex social contexts. Trust over time that the app is reliable and does its job is broadly negative to neutral (Figure 3):

> *I'm not sure that the proximity of devices like smart phones is a good indicator of risk of exposure to an airborne virus.* [Participant E; motivation of use for self-protecting has decreased; change factor of 9]

> *And there were so many cases of people I knew being told to self-isolate who had absolutely no contact…Yeah, you could be next door neighbours with somebody and have no contact with them and still get told to self-isolate and you had no come back on that.* [Participant F; the app's perceived usefulness to self has decreased by 4 points and trust in app reliability and effectiveness decreased by 2 points; has deleted the app; change factor of 18]

> *Without contextual information, people can't make good choices about what the information means. So, the stories that came out very quickly about phones being left outside of lockers in hospitals, and NHS staff being pinged because they didn't turn the app off...You know, people who were getting pinged from across closed walls and things.* [Participant I; initial negative perceptions of the T&T app that were maintained over time; never downloaded it; change factor of 4]

Moreover, in the view of the participants, the existence of these issues has contributed to the general public abandoning the app. Trust over time that most people will download and self-isolate is broadly negative to neutral (Figure 3):

> *If people are not using it, then the T&T app it's like useless.* [Participant A; general positive motivations and attitudes toward the app but decreased belief that people will download it; change factor of 12]

> *It's just technology that is so novel and that it breaks very easily and also people can just choose to not use it, so there's many obstacles for it to be efficient.* [Participant B; decreased motivation of use to protect self and society and decreased belief that others influence own actions and others will download the app; change factor of 25]

1b: Lack of Consideration of User Diversity

The participants expressed that the design of the app did not consider different sets of users and their specific needs and situations. For example, there were sociotechnical flaws arising from the unique conditions of frontline workers, as mentioned by participant F:

> *Because I was every day in school with kids, they weren't testing. We were basically hung out to dry…I think our situation was so different. If I had it [the phone] on all the time in school, I'd have been pinged so many times to self-isolate.* [Participant F; decreased perceived usefulness to self by 4 points and decreased trust that app is reliable and effective by 2 points; has deleted the app; change factor of 18]

Participants also mentioned other sets of users who were not considered by the app designers, such as people sharing workspaces and households and varying levels of smartphone users. Participant I summarized this as follows:

> *And to me the test and trace app was designed with such a Tory model of the world [i.e. this could be interpreted as a conservative and upper-class view], that smartphone users were professionals, that they weren't frontline users, and that they wanted information or would use an app. That the idea that it would be older users with older phones, that it would be people without the latest smartphone, that it would be people who would have a difficulty navigating smart phones. I just felt like there was so many missed opportunities to have a model of a user who didn't fit, who Downing Street [UK government] was clearly designing this thing for.* [Participant I; initial negative perceptions of the T&T app that were maintained over time; never downloaded it; change factor of 4]

1c: T&T App Designed for Individuals Rather Than a Collective Effort

For participants, adherence to the guidance and general uptake of the T&T app as a collective action was needed for effectiveness. However, some participants considered that it was designed for individual action and responsibility instead of creating an app modeled by mutual aid:

> *Having the app puts the responsibility of using it on the individual, it's like it's not a communal or collective solution. It's very much like, does this person have a phone that can support this? Does it have Bluetooth well enough? Does the location tracking work? Is there enough battery for people who have an older phone? Or maybe do not even have a phone at all.* [Participant G; initial mistrust of the T&T app that remained unchanged over time; change factor of 10]

Participants felt that the T&T app did not help society in general, but simply their closest circle, and that it made people lose their goodwill in the general public:

> *I don't really know what that means completely, I feel like I'm helping only my small circle as much as I can, but I don't think I can affect like the broader society.* [Participant B; decreased motivation of use to protect self and society; change factor of 25]

1d: Lack of Clarity and Certainty in Understanding of the T&T App

When prompted to talk about how the T&T app works, there was a lack of certainty in the participants' answers. In the survey, they reported a good understanding of the app and strongly agreed with the need to have explanations and verifications by the T&T app (Figure 3). While some participants mentioned that decisions were made by a combination of the T&T app and humans, other participants mentioned that the decisions were made only by the T&T app. A few participants were in both groups as their answers depended on the specific context discussed (eg, the existence of regional clusters will be handled differently). The participants mentioned their lack of understanding and certainty regarding how the app works and the extent of human involvement in the system:

> *I'm also realising how embarrassingly little I know about how this app works...but my assumption is like almost all automated. I just feel like somewhere in it someone's got the ability to be like let's send a message.* [Participant H; mix of positive and negative responses that remained broadly the same; change factor of 4]

> *You know what? I don't know. I think originally, it was humans and the app, but I think they may have made it more—I'm going to say app only…There are certainly people in the system. But come to think of it now, I think I'm actually very sure they're probably not involved in decisions as to whether or not self-isolate someone.* [Participant C; reported a good understanding of how the app works that was maintained over time; change factor of 8]

1e: Suggestions Given to Improve the T&T App

As the participants discussed which aspects of the T&T app were seen as flawed, they also gave suggestions on what could be improved. Most of the suggestions were related to messaging and communication:

> *I feel like it could have been designed around…positive feeling. Here's a little something "Oh it's cool that you've checked in" or "You checked in five times." Not to gamify everything, that's also really bad, but I think just a little "Thank you for having on contact tracing" that "You've been outside, you are good, remind your family and friends to do that too"...that would have been nice.* [Participant G; maintained motivations and attitudes to protect self and others; change factor of 10]

There were also calls for the T&T app to be more human-centered with simpler and manual check-ins and to help the human T&T team with options to verify the app to avoid scams.

T2: Trust in the T&T App Differs Based on Perceived Usefulness and Functionality

2a: Varying Degrees of Trust

Trust of the participants was primarily based on app functionality and effectiveness. Thus, if the participant believed these factors to be flawed, then trust was low, whereas if the app was thought to be effective, then trust was higher. For example, some participants saw media stories such as the "Pingdemic" (Figure 1) and the fact that real consequences occurred such as shutting down venues as signs that the app was working effectively and thus trust was higher. Whereas other participants thought that the app had poor functionality and effectiveness, which led to a loss of confidence and trust:

> *The "Pingdemic" in my mind just shows the app was doing its job because you know Covid was on the rise and people were getting pinged for it. So, I suppose yes, in that case, I probably do think that it's doing its job.* [Participant C; initial trust in the app that was maintained over time; change factor of 8]

> *Whereas I think if it had worked properly and people had trusted it, then maybe we'd be in a better situation now.* [Participant F; decreased perceived usefulness to self by 4 points and decreased trust that app is reliable and effective by 2 points; has deleted the app; change factor of 18]

> *Your trust is very surrounded by the fact that it works or not.* [Participant I; initial negative perceptions of the T&T app that were maintained over time, never downloaded it; change factor of 4]

2b: Flawed Development and Lack of Explainability and Transparency

This subtheme links to the TAM [24] constructs: perceived usefulness and perceived ease of use. As the T&T app was seen as flawed in both aspects by participants, their acceptance and trust in the app were low. Reasons given for mistrust in the app included flaws such as the T&T app creators' lack of understanding and consideration of all the complex factors needed for the app to be effective and the lack of explanations behind notifications:

> *You know if there could have been any explanation as to why I was being pinged and my husband wasn't. That would have really helped.* [Participant D; decreased perceived usefulness to self and decreased trust in app effectiveness and trustworthiness; has deleted it; change factor of 9]

> *I mean, it doesn't bother me that an outside agency built the app. I just don't think they built it very well...I don't think they knew what they were doing...in their ability to understand the socio-technical complexities of the assemblage that needs to be put in place for this thing to work.* [Participant I; initial negative perceptions of the T&T app that were maintained over time; never downloaded it; change factor of 4]

2c: Human T&T, Physical Measures, and Human Guidance

Participants had a strong view that human T&T and physical measures, such as mask wearing, social distancing, and ventilation, were more trustworthy than the T&T app because the app was fundamentally flawed. In addition to higher trust, there was also a sense of control gained from complying with physical measures, which in turn made participants feel safer:

> *It might be a side effect of this pandemic, that we're all looking perhaps for a little bit more control. But also, it's going to give me a result which I have more confidence in, particularly doing regular testing...Whatever is the result of my weekly tests, that's something which I know I can act upon or deal with, whereas with a ping, all that is, is basically a cause of stress really.* [Participant D; decreased perceived usefulness to self and decreased trust in app effectiveness and trustworthiness; has deleted it; change factor of 9]

Human contact in the T&T app was regarded as important (Figure 3). This was done for multiple reasons, such as ensuring accountability, verifying automated decisions, and avoiding unnecessary actions such as self-isolating:

> *There's an element of it where I think anything that's algorithmic or AI or machine learning...I want some human accountability somewhere in the chain. And that's more of a principle thing. It's kind of less about the app, it's more about like I'm fine with you using AI to improve the service but I still need some accountability and some ability to say, "well, why was this decision reached? How was this decision reached?" And its impact on me. I think a human [needs to be] somewhere in that system.* [Participant H; mix of positive and negative responses that broadly remained the same over time; change factor of 4]

T3: General Low Trust in the Government Owing to Poor Pandemic Handling, Including Procurement of the T&T App

3a: Inconsistent Decision-making and Lack of Compliance

Participants felt anger and frustration toward the UK government because of mixed messaging and the perceived hypocrisy of making rules yet not following them. This, along with perceived poor and inconsistent decision-making throughout the pandemic, led to participants' low trust in the government:

> *I think all the mixed messaging, basically saying "now Covid's your problem, you decide if you want to wear a mask or not," I just think all of that plus an app which is unreliable just makes you not feel particularly trustworthy.* [Participant D; initial negative trust in the UK government that has remained unchanged and decreased trust in the app; change factor of 9]

> *I think over the last year we've had a number of cases of political leaders in England and the UK...who*

appear to have either not been following the rules or have been following them loosely, shall we say, and I think that would have affected trust around these things. [Participant E; initial negative trust in the UK government that has remained unchanged; change factor of 9]

3b: T&T App Creation and Data Management

Another reason for reported mistrust in the UK government was the doubt around the creation of the app and how the data were collected, stored, and processed. Several aspects reported in the media, such as the Excel incident, technical decisions, and money spent, in addition to the UK government's history of failures with technology projects, all resulted in a general mistrust in the government and questioning of their intentions regarding data management:

Because I am a data scientist, so I know what kind of data people use. I just wouldn't want this to be an exercise of the government to collect free data under the excuse that it's for the national security. Yes, I just don't really know who and where this data will be stored considering the problem that happened to that Excel file. It just throws a shadow over anything else. [Participant B; decreased trust in Big Tech; change factor of 25]

The government and big IT projects, I mean, you know there's been a history of failure there, hasn't there? But I just think this was so important to try and get it right and they didn't. [Participant D; initial negative trust in the UK government that has remained unchanged and decreased trust in the app; change factor of 9]

3c: Tensions Between Enforcing the T&T App and Opting Out of Tracing

Within this subtheme, participants expressed morally complex views about their ability to opt in and out of contact tracing. Although participants believed that the freedom to opt-out of contact tracing was important, they also believed that this was a complex decision as the T&T app requires contact tracing to be turned on to be effective. Furthermore, the importance of not only user control within the T&T app (being able to opt-out) but also the need for transparency was highlighted:

It's good for people to have a choice. I mean, yeah, just for people to have a choice to have this contact tracing or not where I ideally—so I think it should be mandatory—so that the virus is not spreading. But at the same time people should have the choice to opt in or out I think. [Participant A; increased belief that opt-out should be an option; change factor 12]

For me the whole point of having the app is for the contact tracing. If I was turning off contact tracing, I might as well uninstall the app. [Participant C; initial disagreement with the idea of opting out of tracing was maintained over time; change factor of 8]

3d: General Disapproval of the Government's Actions During the Pandemic

This subtheme was substantial and included a variety of reasons why participants disapproved of the government's actions over the course of the pandemic. A common opinion among the participants was that the government handled the pandemic poorly and could have implemented the T&T app more effectively. The measures suggested by the government were also thought to be elitist in principle and punished the diverse UK population, including the working class and frontline workers. In addition, the government was considered overconfident, which led to a premature relaxation of rules, with multiple participants reporting that rules and enforcement of rules should have been stricter to improve compliance and reduce case numbers. This was accentuated by some participants' comparison of the England T&T system to pandemic handling in Scotland, which was thought to be clearer and stricter and thus more effective:

I was thinking I didn't agree with how the UK Government handled Covid in general...This is a lay person's opinion, but I would have liked much more conservative approach, let's say closing down earlier. Much more money for people to stay home. Extended furlough. Everybody gets Universal Credit. That's what I would have done but I didn't have to crunch the Excel sheet, so it's easy for me to say that. [Participant G; initial low trust in the government that remained unchanged; change factor of 10]

T4: Varying Degrees of Trust in Stakeholders

4a: Relationship Between Perceived Intentions and Competence

Participants' reported trust in the rest of the stakeholders (ie, private contractors, Big Tech, large and small hospitality venues, local councils, and the NHS) varied in a spectrum of their perceived intentions and competence exhibited throughout the pandemic:

If we've got a spectrum of competence and then good intentions, Serco, Capita et al. private contractors, they're ranking dead last on both criteria. I regard them as both incompetent and malicious. Tech companies I regard as competent and malicious. So, they end up in the middle and then with small hospitality venues, not malicious, like none of the rest really I'd put down as particularly malicious. I'd say that chain restaurants are largely malicious...So for them, I'm measuring on a competence basis...For NHS they get a four because of some worries on competence, the UK Government get a four because of some worries on maliciousness. There's, there's a chart here. [Participant H; mix of trust in stakeholders; change factor of 4]

Building on this categorization stated by participant H, further subthemes were developed in which participants reported mixed feelings regarding trust in stakeholders. These feelings of trust or mistrust depended on stakeholders' perceived intentions and competence, as expressed by the participants (Figures 4 and 5).

It is worth noting that these figures represent the generalized perceptions of participants' views regarding stakeholders and are not intended to be an objective grading or make any assumptions about stakeholders' actual competence or intentions. In addition, these are not equally weighted reflections of the views of all participants involved in the research, as not all participants mentioned all stakeholders, but rather are used here to illustrate participants' general reasoning for trusting—or not trusting—such stakeholders.

Figure 4. Overall reported trust in stakeholders based on their perceived intentions and competence. Figure created based on data from Figure 3. Color key represents average trust across participants (where decimals were rounded to the closest integer) at the 2 time points. Gradient represents change of average trust over time (from left to right), and solid color represents no change in average trust over time across all participants. The size of colored areas aims to illustrate the number of participants that mentioned the stakeholder in question. This figure is meant to illustrate generalized perceptions gathered from interview and survey data and does not represent statistical measurements found in the data. NHS: National Health Service.

Figure 5. Participants' positive and negative views of stakeholders that influence participants' trustworthiness.

High competence Reputation Having adequate resources and support Knowledge and expertise	**Good intentions** Persevering effort and determination, despite lack of resources Encouraging compliance Honesty and consistency
Low competence Poor technology competence or ability Poor decision-making Lack of data security Lack of resources and support	**Bad intentions** Data sharing or misuse Contributing solely for own benefit or profit

4b: Big Tech, Private Contractors, and Large Hospitality Venues

Overall, participants reported mixed feelings regarding Big Tech and large hospitality venues, as their capabilities for managing T&T were regarded as stronger in comparison to the rest of the stakeholders. Some participants mentioned their reputation, expertise, infrastructure, and level of involvement in T&T, all of which provided a general sense of high competence despite the known mistakes and issues arising throughout the pandemic:

> *The bigger you are, if you're a larger venue you might have like a, you know, data control officer...you know, you're large enough to have a corporate policy on it, you know there's at least 50/50 chance that was followed in your particular franchise, so I just think the odds of it being handled appropriately increase once you're a large chain just cause you'll have that corporate infrastructure underpinning it.* [Participant H; neutral to negative trust in large venues; change factor of 4]

Nevertheless, participants expressed general concerns about those main stakeholders (Big Tech, private contractors, and large hospitality venues) for various reasons related to their perceived maliciousness. Big Tech was the primary stakeholder occasioning low trust in this regard. Participants tended to group Big Tech and private contractors as similar entities, and thus, some reported having generally low trust in both. Concerns were linked to the potential misuse of personal data and a perceived interest in getting involved with the T&T app only for profit. Similarly, when comparing small and large hospitality venues, the participants reported lower trust in the latter because of the potential for using T&T data for their own benefit:

> *It just doesn't seem feasible to me that somebody such as a big company would be involved without getting something in return. There must be some sort of profit-making.* [Participant B; decreased trust in Big Tech; change factor of 25]

4c: Small Venues, Local Councils, and the NHS

Conversely, the remaining stakeholders were considered not particularly malicious; however, their specific conditions such as bad outcomes, poor infrastructure, or decentralized procedures were sources of mistrust in them having the capability to adequately handle T&T activities. Participants mentioned trusting the NHS slightly less because of the rising COVID-19 cases. Some participants pointed out the poor treatment of personal data handled at small venues when manually checking in. In fact, most changes in trust to stakeholders occurred in regard to small venues (Figure 3). Similarly, participants recognized that local councils have less budget and poor communication with the centralized government:

> *I would say just from kind of experience that small businesses have a harder time doing data security and data protection work and that larger companies tend to have much bigger infrastructure for doing that.* [Participant I; neutral trust in most stakeholders; change factor of 4]

Despite participants' concerns about how these stakeholders managed T&T and other pandemic-related matters, there was a general feeling of trust in them regarding their intentions. In fact, participants strongly empathized with small hospitality venues, and although they stated that their T&T procedures were not ideal, the harsh conditions they have been through with the pandemic were acknowledged, and thus, their efforts in following guidance were recognized. Similarly, participants considered that local councils made good efforts in handling pandemic issues at their local level and were perceived as more consistent than the central government. Some participants also expressed an existent or renewed general trust and appreciation for the NHS, as it had been put through difficult times:

> *The smaller they are the more sorry I feel for them. They kinda had to do their best. And I remember the time when we still had to sign manually which I assume must have been a nightmare. Especially considering they would have to close down if they didn't have it. So I do trust them...I don't think data was misused.* [Participant B; increased trust in small venues, local councils, and NHS; change factor of 25]

4d: General Mistrust and Uncertainty Regarding Institutions Involved in T&T

Participants reported an overall mistrust of the "whole system," on occasions specifically referring to the central government (as detailed in *T3: General Low Trust in the Government Owing to Poor Pandemic Handling, Including Procurement of the T&T App* section) and at other times conflating Big Tech organizations and other government dependencies. Further, other participants stated a general mistrust of real-world project development, pointing out gaps and malpractices, as well as general concerns regarding personal data treatment regardless of the institution. Nonetheless, these participants acknowledged that some compromises were needed for practical reasons for getting through the pandemic:

> *I wouldn't say it's about trusting the app. I just think the whole system's a bit broken.* [Participant F; decreased trust in small venues and the UK government; change factor of 18]

T5: Over Time, Public Consciousness and Compliance Have Lessened Regarding the Pandemic

5a: Compliance Has Generally Drifted Away Over Time

This subtheme is substantial. Participants reported varying degrees of compliance with the use of the T&T app as well as other social distancing and isolation guidance. The degree of participants' compliance with the guidance was based on internal and external factors. Participants reported following the pandemic measures, including the use of the T&T app to protect others, mentioning a sense of social responsibility. Although participants reported complying with the T&T app despite seeing it as flawed or untrustworthy, they still believed it should be used:

> *My loss of privacy in the bigger scheme of things is not as important for this period of time than potentially protecting the grandma next door. So, I downloaded [the app] despite my concerns.* [Participant G; initial mistrust of data used responsibly and securely that remained unchanged, but high motivation of use to protect self and others; change factor of 10]

On the other hand, when participants reported low compliance, they mentioned a combination of reasons, including the relaxation of the rules and the inconsistency and lack of coherence of guidance. People mentioned still relying on other physical measures such as testing and mask wearing, as well as

a sense of security from being vaccinated. Participants also adapted the rules and guidance according to their own common sense; that is, what they viewed as the most effective or convenient solution for their specific situation:

> *It seems to be a combination of the vaccine, a combination of frustration over the length of the pandemic, or just that sheer, "I've gotta live my life." And, for some friends and families, it's financial pressures.* [Participant E; initial negative trust in app that remained broadly unchanged; change factor of 9]

> *I was every day in a class of 30 kids with no social distancing. It kind of became a bit of what's the point?* [Participant F; decreased perceived usefulness to self by 4 points and decreased trust that app is reliable and effective by 2 points; has deleted the app; change factor of 18]

> *First time I had to use my Covid passport and stuff like that to get into places I was really happy, I just decided you know what, "I'm now in control of this myself, I'm not having the app.*" [Participant D; decreased perceived usefulness to self and decreased trust in app effectiveness and trustworthiness; has deleted it; change factor of 9]

> *I wanted to comply, to self-isolate. And I'd have to cancel the hotel, the trains, and it means a cost for me. So yeah, I did not comply. But we took the test, and we were tested negative, so we proceeded with our holidays.* [Participant A; decreased trust that people will download the app; change factor of 12]]

Interestingly, participants described experiences of compliance fading over time. These experiences referred to participants' own compliance behavior changing, as well as other external factors such as the perceived behavior of the general public. Moreover, participants consistently reported observing venues not complying or prompting customers to follow T&T measures anymore, such as manually entering their details or checking in using the T&T app. They further expressed decreased visibility of QR codes at venues. Some participants believed that this was because T&T came across as not being needed anymore, and others said it reflected how it went out of public consciousness over time (see *5d: Media Exposure Has Decreased* section), although it could still be potentially useful:

> *The QR codes are still going on. I don't think many people are using them. I do try to do them when I remember...and quite honestly, I find myself forgetting.* [Participant C; increased trust that people will download the app and isolate; change factor of 8]

> *It [T&T App] feels less important now it's less visible. It's significantly less visible, and I see many of my friends and family pretty much unaware of that test and trace is still going almost. The other things are because they know the visibility of the whole engine has gone, I still think it could be very important in tracking down outbreaks or in spotting new variants.* [Participant E; decreased motivation of use to protect self; change factor of 9]]

5b: Early Interest and Curiosity in the T&T App Has Not Sustained

The participants also expressed dedicating some time to investigating on the web about the T&T app when it was first announced and released and were mainly curious about T&T app functionality. Some participants were concerned about data protection and thus closely followed media reports about T&T app development. However, participants also acknowledged that they stopped reading about the app, both because there were not as many media reports as at the beginning of the pandemic and because some eventually lowered their use of the T&T app:

> *Early doors I did, so when it first came out and the conversations were around, cause they partnered with like Google and a couple of other tech companies, and I was interested in the conversations about how much personal data was in there, how would be used, well like protections…So I was interested at that point, I can't say I've read anything about the Covid-19 app in the last six months.* [Participant H; mix of positive and negative responses that broadly remained unchanged; change factor of 4]

5c: Changed and Sustained Feelings Regarding T&T App

Finally, this subtheme was developed to reflect the participants' changed and sustained feelings toward the T&T app. Participants who initially had negative expectations or assumptions about the app maintained those negative perceptions over time because of the government handling of the pandemic (described in *T3: General Low Trust in the Government Owing to Poor Pandemic Handling, Including Procurement of the T&T App* section) and the drop in T&T app use (described in *5a: Compliance Has Generally Drifted Away Over Time* section):

> *I was very sceptical and nothing has changed to make me less sceptical. I was very sceptical that it would work and you know, I'm, I consider myself very socially minded so I do believe in the good of society. I would care about the good of society, right? But I didn't see how the test and trace app was actually for the good of society.* [Participant I; initial negative perceptions of the T&T app that were maintained over time; never downloaded it; change factor of 4]

Moreover, only one participant accounted for their positively changed trust in the T&T app over time, which was prompted by their questionnaire answers from a previous study:

> *I think a lot of it is because I took that survey really early on in getting the app and at the time I do think there were more problems with it. Bugging out or crashing or being a bit crap. And it's kind of proven itself and had a couple of improvements overtime. So I've probably yeah, I've probably grown to think the app is better than I did when it first launched for it seemed a bit more slapdash.* [Participant H; increased trust in app reliability; change factor of 4]

5d: Media Exposure Has Decreased

Although some participants reported media content adding to people's negative perception of the T&T app, over the last year, the media content regarding T&T has decreased in favor of other topics considered more sensational. Some participants believed that the media and the government should increase exposure of the T&T app so that people are reminded of its existence and thus are prompted to use it:

> *I think that then it says on the UK Government to talk less about it. So it makes me, and maybe people, less concerned about it. So maybe if you just keep it in the same level. I mean, keep informed, persuade us to use the T&T app or to still doing the measures. They talk less about it, I think.* [Participant A; decreased trust in small and large venues, positive unchanged trust in the UK government; change factor of 12]

> *Didn't think it was such a newsworthy thing anymore because there was, I mean pick, take your pick in all the things, in the scandals and what people do. And it is also not very juicy reporting to kinda "Oh we're still using the app," "keep using app"…There's much more juicy [stuff] to read about so, I think it just became boring probably.* [Participant G; initial negative trust in most stakeholders that remained unchanged; change factor of 10]

Discussion

Principal Findings

A total of 5 themes were developed from the qualitative interview data, finding multiple reasons for changes in use of the T&T app. These are also reflected on the varied change factor in survey responses over time, the lowest being 4 points and the highest being 25 points. The largest contributors to change over time were the flaws experienced when using the T&T app (T1) and the lack of trust in the UK government because of how the COVID-19 pandemic was handled (T3). Other factors influencing trust included perceived usefulness and functionality of the app (T2), trust in stakeholders (T4), and public consciousness and compliance lessening over time (T5).

The results of this study elucidate on the concept of sustained use in the context of DCT, which for the T&T app consisted of a range of actions, some more passive than others; for instance, keeping Bluetooth tracing on or more actively scanning the QR codes at public venues. In this paper, we move beyond the TAM model [24] and instead adopt the concepts proposed by the TAL model (ie, the transition from preuse acceptance to initial and sustained use) [26] to explain participants' experiences with the T&T app over time, which are influenced by their perceived usefulness and ease of use among other factors such as trust in the stakeholders involved. After a year of use, the NE of the T&T app wore off, as expressed in T5, in line with the literature [28]. Personal and social motivations to use the app have also changed (Figure 2), which are some of the factors that lead to sustained use of technology [29]; however, a few participants continued using it as it was a requirement for their job.

Trust is a complex topic but was generally reported to be influenced by a combination of perceived intentions and competence. Over this year, participants' trust in involved stakeholders has also slightly changed according to Figure 3—trust in government (partly because of the scandals discussed in the introduction), large venues, and Big Tech decreased; trust in local councils and NHS increased; and trust in small venues increased for some participants and decreased for others (T3 and T4). There were some slight differences between the trust scores given in the surveys and the findings from the interviews. Furthermore, when the survey trust scores were averaged from both time points, most changes were not substantial (Figures 3 and 4), which further establishes the variable and subjective nature of trust. Trust in the stakeholders that form the app's ecosystem influences trust in the app and its uptake [22,23]. The combination of the NE "wearing off," lack of personal and social motivation for app uptake, and general low trust in T&T app stakeholders were reasons given by participants explaining their change in the use of the T&T app, as markedly evidenced by 2 of the participants deleting the app between the initial survey and the interview.

Studies investigating attitudes toward DCT apps have identified that people were positive about and intended to use them to help mitigate the spread of COVID-19 and protect others [3,11], which is broadly reflected by the statements provided by the participants of this study (Figure 3). Nonetheless, our results suggest that although social influences can be a motivator for adoption [7,10,16], changes in the use of the T&T app were occasioned by several factors such as experienced and perceived flaws, mistrust surrounding the whole app ecosystem, and everyday life practicalities and contingencies. Then, this study both confirms the intention-behavior gap identified in previous studies of DCT [11] and contributes to providing some of the reasons for its occurrence.

In line with the fifth theme developed in the thematic analysis, media representation and concern regarding the pandemic lessened over time, which appeared to have a direct effect on the behavior of participants. Although the initial intention to use the T&T app was positive as discussed earlier, the normalization of the pandemic in the media, along with a growing sense of pandemic fatigue, led to decreased use or deletion of the T&T app. Normalization and pandemic fatigue were therefore 2 key factors that had an impact on the compliance and behavior surrounding the use of the T&T app, despite the initial intention from participants to continue using the app. A further explanation for the lack of trust and poorly sustained use of the T&T app could be a "learned helplessness" developing in individuals because of consistent failures from both the UK government and from the technological capabilities of the T&T app. "Learned helplessness" is a learned state that develops from powerlessness arising from uncontrollable traumatic events, leading to the general belief that a situation is unchangeable [40]. As the COVID-19 pandemic was out of anyone's individual control, and efforts to reduce the spread of the virus were appearing unsuccessful owing to the rise in cases, it is possible that people began to feel a sense of learned helplessness, which in turn led to complacency with using the T&T app. These explanations are consistent with previous

findings, which highlight that a decrease in concern, low trust in political systems, and complacency can negatively affect the adoption of DCT apps [41,42].

Measures to stop the spread of COVID-19, like the uptake of DCT, are of a collective nature owing to the behavior of the virus. This tension between the need for a collective response and the individual-based design of the T&T app is shown in subtheme 1c. Fischer [43] demonstrated that the individual-collective nature of a society influences its collective actions regarding COVID-19 behavior, where more economically advantaged and individualistic societies have weaker collective action properties such as this study's context (United Kingdom). Thus, the cultural context of the United Kingdom could be another factor influencing the intention-behavior gap identified in this study.

Finally, as some have started to point out [16], the results of this study divert from previous work reporting privacy and security concerns as major barriers to the adoption and use of DCT [3,4,44-47]. Although some participants stated having such worries, the perceived benefit of DCT overtook them. This occurrence may also be explained by the normalization of affective discomfort [48], by which people continue using apps despite considering them as dubious. Hence, although privacy and security may play an important role in the initial adoption and use, in the long term, these concerns moved to the background for our participants, possibly facilitated by a lack of major data breaches taking place. Furthermore, this study expands on the reasoning for mistrust in governments deploying DCT, beyond worries of massive surveillance [3,46]; as elaborated in T3, it is also constructed by people's assessment of the government's capabilities for managing the pandemic and creating and managing the T&T app.

Practical Implications

Although this study operates in the specific context of the United Kingdom, several implications and lessons can be learned from the individual and collective experiences of people with the T&T app after a year of deployment. First, participants encountered by themselves, or as well-known social occurrences, a number of flaws with the app ranging from technical issues to little consideration of user diversity and how the app would be used in different situations. Moreover, the participants expressed a lack of clarity and certainty in their understanding of how the app works. Therefore, although people have good intentions to support society (Figure 3), or even if the apps are marketed to appeal to good citizenship and collectivism [7,10,49], our study suggests that if people do not see how DCT is achieving such a goal, sustained use becomes hindered. Alternatively, people who continue using the app do it despite not being sure if their actions are contributing to controlling the pandemic. Although this research suggests such implications, it is important to maintain that these conclusions were gathered from a small sample and cannot therefore be widely generalized.

The implication for future DCT systems is that besides considering a range of real-world scenarios (eg, multioccupant or shared-wall households) and a diverse set of users (eg, frontline workers), they must provide further contextual information to explain to users how decisions are being made by DCT apps and ensure transparency of the technical (eg, false positives) and practical (eg, effectiveness at large) matters. Moreover, this study shows that the deployment of DCT apps should go hand-in-hand with other measures to avoid provoking perceptions of uselessness. At a very minimum, DCT apps should keep being promoted over time by the organizations involved in deploying them. Other steps to improve sustained use could be taken by exploring the design of DCT that addresses the loss of NE and how trust in the whole ecosystem (app, organizations, and other users) can be strengthened.

Finally, participants' accounts of their experiences using—or not using—the T&T app beg the question whether DCT is effective or needed at all [50]. This study aligns with findings from Tretiakov and Hunter [16], in which DCT actual use declined when the alert or risk levels were low. As we move into the endemic stage of COVID-19, the practical application of proximity-based DCT needs to be reassessed and must work in combination with other physical measures, such as vaccines and testing, that give people more reassurance and clearer results upon which they can act. Some directions include the development of hybrid contact tracing systems that integrate the participation of human contact tracers in the whole ecosystem [51].

Limitations

There are some limitations of this research that should be highlighted. First, participants in the interview study were somewhat homogeneous demographically and could be considered a small sample size. This research only gathered qualitative data from 1 frontline worker (for as much as we know from the interview discussions, in which affected occupations were likely to be discussed), and all the participants had received higher education degrees (Table 1). Therefore, this sample and the subsequent thematic analysis may not be representative of a diverse population. In addition, Dowthwaite et al [19] identified statistically different responses for Black, Asian, and minority ethnic participants in their survey study. This could not be investigated in this study because only 1 out of 9 participants identified as member of the Black, Asian, and minority ethnic group; thus, these differences should be explored in future research.

Moreover, although participants were asked about the impact of the UK COVID-19 vaccination rollout on trust, the data on this were not substantial enough to form a robust theme. The impact of vaccinations was only mentioned by 1 participant in theme 5. This is not to say that vaccination rollout did not influence how people perceived the dangers of COVID-19 and the role of preventative measures, just that the individuals in this study did not focus on the topic when discussing their views on the T&T app and the factors influencing trust. It could be argued that many participants in this research were angered and frustrated by the UK government's poor handling of the pandemic, which overpowered the positive influences such as the vaccination rollout in the United Kingdom.

Another limitation of this research is the lack of justification for the questionnaire scores at time point 1. As only quantitative data were gathered at the first time point, the authors could only truly compare the quantitative data longitudinally. The thematic

analysis should only be considered longitudinal in a retrospective manner for descriptive explanations (participants were asked to reflect on how their trust and general views of the T&T app had changed over time). Consequently, this may have resulted in the possibility of recall bias and inaccurate portrayals of experience.

Finally, this study is geographically limited to the specific UK context, in which a centralized government and public health system exist. Thus, it could be argued that the findings of this research are only transferable to places with similar political, cultural, and health systems, if at all.

Conclusions

To conclude, this research aimed to understand how the use of the T&T app and trust changed over time. By conducting interviews and exploring survey answers approximately 1 year apart, we found multiple reasons for changes in trust and diminishing use. For instance, the 2 largest contributors to change were the perceived flaws in the T&T app and a lack of trust in the UK government owing to the way the pandemic was handled. In addition, multiple factors impacted the participants' compliance with the app. Initial NEs occurred with the T&T app, which lessened over time as a concern and media representation of the pandemic decreased and a new norm was established. These findings are an important initial step for future technology and app design and to increase understanding around how the general public perceives and trusts in the technology used for health care, and which factors influence the uptake and sustained use of DCT apps.

Acknowledgments

This research was supported by the Engineering and Physical Sciences Research Council (grants EP/V00784X/1, EP/M02315X/1, and EP/T022493/1). The survey was refined based on feedback by Ipsos MORI. The authors also thank the reviewers for their insightful feedback that has resulted in an improved paper. CMB acknowledges the financial support of National Institute of Health and Care Research MindTech MedTech Co-operative and National Institute of Health and Care Research Nottingham Biomedical Research Centre.

Authors' Contributions

LD and JEF were involved in the conceptualization and methodology of the study. LD conducted the interviews, administered the questionnaire, conducted analysis of the quantitative data, and produced supporting figures. CP, GRC, and ARP equally conducted the thematic analysis and wrote the original draft. CMB and HW advised and supported CP, GRC, and ARP on the thematic analysis. JEF contributed toward funding acquisition and project administration. JEF, LD, and EN contributed to reviewing and editing of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statements presented in the web-based questionnaire and revisited in the interviews.
[DOCX File , 20 KB - jmir_v24i10e40558_app1.docx]

Multimedia Appendix 2
Participants' details, attitudes, and changes from the initial questionnaire to interview.
[DOCX File , 17 KB - jmir_v24i10e40558_app2.docx]

Multimedia Appendix 3
List of themes and subthemes generated through thematic analysis.
[DOCX File , 18 KB - jmir_v24i10e40558_app3.docx]

References

1. Du L, Raposo VL, Wang M. COVID-19 contact tracing apps: a technologic tower of babel and the gap for international pandemic control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194. [doi: 10.2196/23194]
2. Blasimme A, Ferretti A, Vayena E. Digital contact tracing against COVID-19 in Europe: current features and ongoing developments. Front Digit Health 2021 Jun 17;3:660823 [FREE Full text] [doi: 10.3389/fdgth.2021.660823] [Medline: 34713135]
3. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]

4. Sharma S, Singh G, Sharma R, Jones P, Kraus S, Dwivedi YK. Digital health innovation: exploring adoption of COVID-19 digital contact tracing apps. IEEE Trans Eng Manage 2020:1-17. [doi: 10.1109/tem.2020.3019033]
5. Utz C, Becker S, Schnitzler T, Farke F, Herbert F, Schaewitz L, et al. Apps against the spread: privacy implications and user acceptance of COVID-19-related smartphone apps on three continents. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445517]
6. Huang Y, Obada-Obieh B, Redmiles E, Lokam S, Beznosov K. COVID-19 Information-Tracking solutions: a qualitative investigation of the factors influencing people's adoption intention. In: Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval. 2022 Presented at: CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval; Mar 14 - 18, 2022; Regensburg Germany. [doi: 10.1145/3498366.3505756]
7. Seberger J, Patil S. Us and them (and it): social orientation, privacy concerns, and expected use of pandemic-tracking apps in the United States. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445485]
8. Fox G, Clohessy T, van der Werff L, Rosati P, Lynn T. Exploring the competing influences of privacy concerns and positive beliefs on citizen acceptance of contact tracing mobile applications. Comput Human Behav 2021 Aug;121:106806. [doi: 10.1016/j.chb.2021.106806]
9. Kaptchuk G, Goldstein DG, Hargittai E, Hofman JM, Redmiles EM. How good is good enough? Quantifying the impact of benefits, accuracy, and privacy on willingness to adopt COVID-19 decision aids. Digital Threats 2022 Sep 30;3(3):1-18. [doi: 10.1145/3488307]
10. Dooley S, Turjeman D, Dickerson J, Redmiles E. Field evidence of the effects of privacy, data transparency, and pro-social appeals on COVID-19 app attractiveness. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; Apr 29-May 5, 2022; New Orleans LA USA. [doi: 10.1145/3491102.3501869]
11. Villius Zetterholm M, Lin Y, Jokela P. Digital contact tracing applications during COVID-19: a scoping review about public acceptance. Informatics 2021 Jul 22;8(3):48. [doi: 10.3390/informatics8030048]
12. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621. [doi: 10.1016/s2589-7500(20)30184-9]
13. Grekousis G, Liu Y. Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review. Sustain Cities Soc 2021 Aug;71:102995 [FREE Full text] [doi: 10.1016/j.scs.2021.102995] [Medline: 34002124]
14. Moreno López JA, Arregui García B, Bentkowski P, Bioglio L, Pinotti F, Boëlle PY, et al. Anatomy of digital contact tracing: role of age, transmission setting, adoption, and case detection. Sci Adv 2021 Apr;7(15):eabd8750 [FREE Full text] [doi: 10.1126/sciadv.abd8750] [Medline: 33712416]
15. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature 2021 Jun 12;594(7863):408-412. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]
16. Tretiakov A, Hunter I. User experiences of the NZ COVID tracer app in New Zealand: thematic analysis of interviews. JMIR Mhealth Uhealth 2021 Sep 08;9(9):e26318 [FREE Full text] [doi: 10.2196/26318] [Medline: 34292868]
17. Arzt S, Poller A, Vallejo G. Tracing contacts with mobile phones to curb the pandemic: topics and stances in people's online comments about the official German contact-tracing app. In: Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411763.3451631]
18. Jamieson J, Epstein D, Chen Y, Yamashita N. Unpacking intention and behavior: explaining contact tracing app adoption and hesitancy in the United States. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; Apr 29-May 5, 2022; New Orleans LA USA. [doi: 10.1145/3491102.3501963]
19. Dowthwaite L, Fischer J, Perez Vallejos E, Portillo V, Nichele E, Goulden M, et al. Public adoption of and trust in the NHS COVID-19 contact tracing app in the United Kingdom: quantitative online survey study. J Med Internet Res 2021 Sep 17;23(9):e29085 [FREE Full text] [doi: 10.2196/29085] [Medline: 34406960]
20. Samuel G, Lucivero F, Johnson S, Diedericks H. Ecologies of public trust: the NHS COVID-19 contact tracing app. J Bioeth Inq 2021 Dec 05;18(4):595-608 [FREE Full text] [doi: 10.1007/s11673-021-10127-x] [Medline: 34609676]
21. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
22. Cheshire C. Online trust, trustworthiness, or assurance? Daedalus 2011 Oct;140(4):49-58. [doi: 10.1162/daed_a_00114] [Medline: 22167913]
23. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]

24. Silva P. Davis' Technology Acceptance Model (TAM) (1989). In: Information Seeking Behavior and Technology Adoption: Theories and Trends. Pennsylvania, United States: IGI Global; 2015.
25. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
26. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. J Med Internet Res 2020 Jul 06;22(7):e17256 [FREE Full text] [doi: 10.2196/17256] [Medline: 32628122]
27. Sung J, Christensen H, Grinter R. Robots in the wild: understanding long-term use. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction. 2009 Presented at: HRI09: International Conference on Human Robot Interaction; Mar 9 - 13, 2009; La Jolla California USA. [doi: 10.1145/1514095.1514106]
28. Mutsuddi A, Connelly K. Text messages for encouraging physical activity are they effective after the novelty effect wears off? In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2012; San Diego, California, United States. [doi: 10.4108/icst.pervasivehealth.2012.248715]
29. Shin G, Feng Y, Jarrahi MH, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]
30. Ceci L. Cumulative downloads of the NHS COVID-19 app in England and Wales as of December 2021. Statista. 2022 Jan 7. URL: https://www.statista.com/statistics/1190062/covid-19-app-downloads-uk/ [accessed 2022-09-13]
31. Timeline of UK government coronavirus lockdowns and restrictions. Institute for Government. URL: https://www.instituteforgovernment.org.uk/sites/default/files/timeline-coronavirus-lockdown-december-2021.pdf [accessed 2022-06-24]
32. NHS COVID-19 app launches across England and Wales. Department of Health and Social Care. 2020 Sep 24. URL: https://www.gov.uk/government/news/nhs-covid-19-app-launches-across-england-and-wales [accessed 2022-06-24]
33. Timmins N. Schools and Coronavirus: the government's handling of education during the pandemic Internet. Institute for Government. 2021 Aug. URL: https://www.instituteforgovernment.org.uk/sites/default/files/publications/schools-and-coronavirus.pdf [accessed 2022-06-24]
34. Sasse T. Coronavirus vaccine rollout. Institute for Government. 2020 Dec 22. URL: https://www.instituteforgovernment.org.uk/explainers/coronavirus-vaccine-rollout [accessed 2022-08-09]
35. Dominic Cummings: what is the scandal about? BBC News. 2020 May 26. URL: https://www.bbc.co.uk/news/uk-52811168 [accessed 2022-06-24]
36. Kelion L. Excel: why using Microsoft's tool caused Covid-19 results to be lost. BBC News. 2020 Oct 5. URL: https://www.bbc.com/news/technology-54423988 [accessed 2022-06-14]
37. Rimmer A. Sixty seconds on . . . the pingdemic. BMJ 2021 Jul 19;374:n1822. [doi: 10.1136/bmj.n1822] [Medline: 34281909]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, California: SAGE Publications; 2021.
40. Seligman ME. Learned helplessness. Annu Rev Med 1972 Feb;23(1):407-412. [doi: 10.1146/annurev.me.23.020172.002203] [Medline: 4566487]
41. Lalot F, Heering MS, Rullo M, Travaglino GA, Abrams D. The dangers of distrustful complacency: Low concern and low political trust combine to undermine compliance with governmental restrictions in the emerging Covid-19 pandemic. Group Process Intergroup Relat 2020 Oct 30;25(1):106-121. [doi: 10.1177/1368430220967986]
42. Hogan K, Macedo B, Macha V, Barman A, Jiang X. Contact tracing apps: lessons learned on privacy, autonomy, and the need for detailed and thoughtful implementation. JMIR Med Inform 2021 Jul 19;9(7):e27449 [FREE Full text] [doi: 10.2196/27449] [Medline: 34254937]
43. Fischer R, Karl J. Predicting behavioral intentions to prevent or mitigate COVID-19: a cross-cultural meta-analysis of attitudes, norms, and perceived behavioral control effects. Soc Psychol Personal Sci 2021 Jun 03;13(1):264-276. [doi: 10.1177/19485506211019844]
44. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
45. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
46. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2021 Aug 16;190(3):863-887 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
47. Duan SX, Deng H. Hybrid analysis for understanding contact tracing apps adoption. Ind Manag Data Syst 2021 Apr 30;121(7):1599-1616. [doi: 10.1108/imds-12-2020-0697]
48. Seberger J, Shklovski I, Swiatek E, Patil S. Still creepy after all these years: the normalization of affective discomfort in app use. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI

'22: CHI Conference on Human Factors in Computing Systems; Apr 29 - May 5, 2022; New Orleans LA USA. [doi: 10.1145/3491102.3502112]
49. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. Eur J Inf Syst 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
50. Maccari L, Cagno V. Do we need a contact tracing app? Comput Commun 2021 Jan 15;166:9-18 [FREE Full text] [doi: 10.1016/j.comcom.2020.11.007] [Medline: 33235399]
51. Lu XL, Reynolds T, Jo E, Hong H, Page X, Chen Y, et al. Comparing perspectives around human and technology support for contact tracing. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445669]

Abbreviations

DCT: digital contact tracing
NE: novelty effect
NHS: National Health Service
T&T: Test & Trace
TAL: technology acceptance life cycle
TAM: technology acceptance model

Edited by C Basch; submitted 27.06.22; peer-reviewed by T Reynolds, V von Wyl; comments to author 27.07.22; revised version received 15.08.22; accepted 16.09.22; published 14.10.22.

Please cite as:
Pepper C, Reyes-Cruz G, Pena AR, Dowthwaite L, Babbage CM, Wagner H, Nichele E, Fischer JE
Understanding Trust and Changes in Use After a Year With the NHS COVID-19 Contact Tracing App in the United Kingdom: Longitudinal Mixed Methods Study
J Med Internet Res 2022;24(10):e40558
URL: https://www.jmir.org/2022/10/e40558
doi:10.2196/40558
PMID:36112732

Original Paper

Characterizing Thrombotic Complication Risk Factors Associated With COVID-19 via Heterogeneous Patient Data: Retrospective Observational Study

Bedda Rosario[1*], PhD; Andrew Zhang[2*]; Mehool Patel[3], MD; Amol Rajmane[4], MD, MBA; Ning Xie[2], MSc; Dilhan Weeraratne[5*], PhD; Gil Alterovitz[2*], PhD

[1]IBM, Round Rock, TX, United States
[2]Biomedical Cybernetics Laboratory, Brigham and Women's Hospital, Boston, MA, United States
[3]IBM, Akron, OH, United States
[4]IBM, Palo Alto, CA, United States
[5]IBM, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Gil Alterovitz, PhD
Biomedical Cybernetics Laboratory
Brigham and Women's Hospital
75 Francis Street
Boston, MA, 02115
United States
Phone: 1 617 329 1445
Email: ga@alum.mit.edu

Abstract

Background: COVID-19 has been observed to be associated with venous and arterial thrombosis. The inflammatory disease prolongs hospitalization, and preexisting comorbidities can intensity the thrombotic burden in patients with COVID-19. However, venous thromboembolism, arterial thrombosis, and other vascular complications may go unnoticed in critical care settings. Early risk stratification is paramount in the COVID-19 patient population for proactive monitoring of thrombotic complications.

Objective: The aim of this exploratory research was to characterize thrombotic complication risk factors associated with COVID-19 using information from electronic health record (EHR) and insurance claims databases. The goal is to develop an approach for analysis using real-world data evidence that can be generalized to characterize thrombotic complications and additional conditions in other clinical settings as well, such as pneumonia or acute respiratory distress syndrome in COVID-19 patients or in the intensive care unit.

Methods: We extracted deidentified patient data from the insurance claims database IBM MarketScan, and formulated hypotheses on thrombotic complications in patients with COVID-19 with respect to patient demographic and clinical factors using logistic regression. The hypotheses were then verified with analysis of deidentified patient data from the Research Patient Data Registry (RPDR) Mass General Brigham (MGB) patient EHR database. Data were analyzed according to odds ratios, 95% CIs, and *P* values.

Results: The analysis identified significant predictors ($P<.001$) for thrombotic complications in 184,831 COVID-19 patients out of the millions of records from IBM MarketScan and the MGB RPDR. With respect to age groups, patients 60 years and older had higher odds (4.866 in MarketScan and 6.357 in RPDR) to have thrombotic complications than those under 60 years old. In terms of gender, men were more likely (odds ratio of 1.245 in MarketScan and 1.693 in RPDR) to have thrombotic complications than women. Among the preexisting comorbidities, patients with heart disease, cerebrovascular diseases, hypertension, and personal history of thrombosis all had significantly higher odds of developing a thrombotic complication. Cancer and obesity were also associated with odds>1. The results from RPDR validated the IBM MarketScan findings, as they were largely consistent and afford mutual enrichment.

Conclusions: The analysis approach adopted in this study can work across heterogeneous databases from diverse organizations and thus facilitates collaboration. Searching through millions of patient records, the analysis helped to identify factors influencing

a phenotype. Use of thrombotic complications in COVID-19 patients represents only a case study; however, the same design can be used across other disease areas by extracting corresponding disease-specific patient data from available databases.

(J Med Internet Res 2022;24(10):e35860) doi:10.2196/35860

KEYWORDS

COVID-19; thrombotic complications; logistic regression; EHR; electronic health record; insurance claims data

Introduction

The World Health Organization reported over 270 million positive cases for COVID-19 and over 5.3 million deaths from the virus worldwide as of December 14, 2021 [1]. As infected patients demonstrate vastly different outcomes, it is critical to identify key patient characteristics that govern the course of the disease across large patient cohorts as early as possible to help allocate the right resources and improve patient outcomes [2]. Logistic regression and machine-learning algorithms have been used to predict which COVID-19 patients will require hospitalization and intensive care to ensure that resources are prioritized to individuals with the highest risk [3-7]. Many of these algorithms make use of routinely collected clinical data.

Although it is well-established that COVID-19 is associated with respiratory complications, the disease has also been observed to cause venous and arterial thrombosis [8]. A hyperinflammatory response has been associated with COVID-19 in increasing the risk of thrombosis [9]. The inflammatory disease process, prolonged hospitalization, and preexisting comorbidities can all contribute to the aggressive thrombotic burden in patients with thrombosis [10-13]. A study in two Dutch university hospitals and one Dutch teaching hospital showed a 31% incidence of thrombotic complications in patients in the intensive care unit (ICU) with COVID-19 [14]. Similarly, the incidence of venous thromboembolism (VTE) in ICU patients was reported to be 25% at Union Hospital, Wuhan, China [15]. In general, VTE has been found to affect up to 46% of hospitalized patients with COVID-19 [16], and a meta-analysis suggested that COVID-19 patients with thrombotic complications have a 2.1-fold higher risk of mortality than those without thrombotic complications [17]. However, VTE and other related vascular complications may go unnoticed in critical care settings [18,19]. As such, early risk stratification is clinically critical for the COVID-19 patient population [20].

There are several potential hypotheses on the mechanisms that may be associated with or responsible for thrombotic complications. For example, there is some preliminary evidence that autoimmune reactions may play a role [21]. In addition, drug interactions are treatment challenges introduced by the therapeutic agents available for COVID-19 [22]. As the population of patients recovering from COVID-19 is steadily growing, a systematic study of the sequelae during the postacute COVID-19 phase is important to collect clinical and scientific evidence to determine the best care for these patients. Furthermore, thromboembolic complications have been reported as a part of postacute COVID-19 syndrome [23-25]. Accordingly, the aim of this study was to use real-world data evidence toward building the foundation for development of a software system that systematically identifies factors affecting VTE in COVID-19 patients.

Electronic health records (EHRs) are widely becoming adopted in health care systems with increasing capability of record sharing across different organizations [22]; however, there remain constraints in using such data along with challenges in gaining unrestricted access. Insurance claims data capture information from all doctors and providers, whereas EHR data capture only the portion of care provided by doctors using the EHR. However, insurance claims data also have limitations such as that these data only cover insured patients. We aimed to bridge these gaps between EHR and claims data, and accommodate both data sources to take advantage of a wider range of data. This design was particularly useful to synthesize a hypothesis regarding COVID-19 and thrombotic complications from IBM and Mass General Brigham (MGB; Boston, Massachusetts) data using IBM's MarketScan claims data set, which was cross-verified with MGB's EHR-derived database. This analysis thus provided a useful approach to bridge the gap with EHR data sets without requiring Health Insurance Portability and Accountability Act–level individual patient information, thereby avoiding the multiple-step process for access.

During the global COVID-19 pandemic, collaboration between organizations has accelerated understanding of the SARS-CoV-2 virus and the COVID-19 disease it causes. While EHR data are widely used in COVID-19 retrospective studies [6,22-26], some organizations use proprietary databases. We have been working to design a method that can handle different types of health care data storage, including the standardized EHR databases as well as any other proprietary data sources such as the insurance claims database used in this study. To respect patient privacy concerns, we only used deidentified patient data when querying the databases. These measures were chosen to make it easier for the work to potentially be used for global collaboration in the COVID-19 pandemic and other cases. This research characterizes thrombotic complication risk factors associated with COVID-19 using information from EHR and insurance claims databases. Comprehensive treatment guidelines and reviews can be found in prior literature [27,28].

Methods

Data Collection

This retrospective observational study utilized deidentified data from IBM's MarketScan commercial claims database. These data were compared and validated with data from the MGB EHR. Adult patients with a COVID-19 diagnosis between February 1, 2020, and September 30, 2020, were included in the study. Patient demographics included age, gender, ethnicity

(EHR database only), and geographic location. We focused on the following comorbidities: hypertensive disease, diabetes, cancer, respiratory diseases (asthma, acute respiratory distress syndrome, chronic bronchitis, emphysema, bronchiectasis, and chronic obstructive pulmonary disease), heart disease (coronary artery disease, heart failure, cardiomyopathy, atrial fibrillation, and ischemic heart disease), cerebrovascular disease (stroke and cerebrovascular disease), liver disease, kidney disease, prior history of thrombosis, HIV, pregnancy, sleep apnea, tobacco smoking use, and obesity. Interventions included veno-venous extracorporeal membrane oxygenation (ECMO), mechanical ventilation, extraneous oxygen use, and medications. The thrombotic complications focused on ST elevation myocardial infarction (STEMI) and non-STEMI myocardial infarction, pulmonary embolism, cerebral infarction, arterial embolism and thrombosis, other venous embolism and thrombosis, transient ischemic attacks and related syndromes, other acute ischemic heart diseases, and other cerebrovascular diseases.

Mapping of Diagnosis Codes

This study included patients with a confirmed COVID-19 diagnosis (International Classification of Diseases, Tenth Revision [1CD-10] diagnosis codes U071, B342, Z8616, J1282, B9729) between February 1, 2020, and September 30, 2020. The outcome of interest was a thrombosis diagnosis (ICD-10 diagnosis codes I21, I24, I26, I63, I74, I82, Z8671, M622, and G45) between February 1, 2020, and September 30, 2020.

Querying Data From the Claims Database

We performed a retrospective analysis of the IBM MarketScan Commercial Database and Medicare Supplemental Database from February 1, 2020, to September 30, 2020, to identify patients. This represents the most recently available data at the time of analysis in IBM MarketScan Treatment Pathways, a cloud-based analytic interface that overlays onto MarketScan Research Databases. MarketScan is one of the largest deidentified longitudinal patient-level health databases in the United States, which includes information on over 39 million individuals, including active employees and their dependents, early retirees, and Consolidated Omnibus Budget Reconciliation Act (COBRA) continuers, insured by approximately 40 employer-sponsored health plans representing all 50 states. A total of 259,470 patients had received a COVID-19 diagnosis at some point between February 1, 2020, and September 30, 2020. Of these, 153,137 patients were continuously enrolled for 2 years prior to the COVID-19 diagnosis and were included in the study.

As an insurance claims database, MarketScan encompasses information from multiple providers in the patient journey with a broader nationwide reach. Insurance claims data provide information on whether a prescription was filled, as opposed to EHR data that only state whether or not a drug was prescribed. MarketScan can effectively complement EHR data by providing an extremely broad view of a patient's interactions across the continuum of the health care system and by providing access to large and diverse samples.

It should be noted that a few of the individuals may drop in and out of the MarketScan data set due to health insurance coverage changes. Hence, while performing these analyses using MarketScan (or any other claims data set), samples are restricted to patients who are continuously enrolled over the observation period.

Querying Data From the EHR Database

We gathered patient data from the MGB patient record database Research Patient Data Registry (RPDR), a centralized clinical data registry. The data warehouse includes 6.5 million patients and 2.2 billion rows of clinical data, serving as a central clinical data registry for inpatient and outpatient encounters from various hospital systems to support clinical research.

The RPDR query tool allows for a search for the number of patients at the hospital with a given set of characteristics. We searched for patients at the hospital between February 1, 2020, and September 30, 2020. Patients were characterized using ICD-10 medical codes for the respective medical conditions with a combination of codes to identify COVID-19 patients with thrombosis and potential associated comorbidities. A total of 31,364 patients had received a COVID-19 diagnosis from February 1, 2020, to September 30, 2020, and were included in the study.

Drawing and Verifying Hypotheses Using Logistic Regression

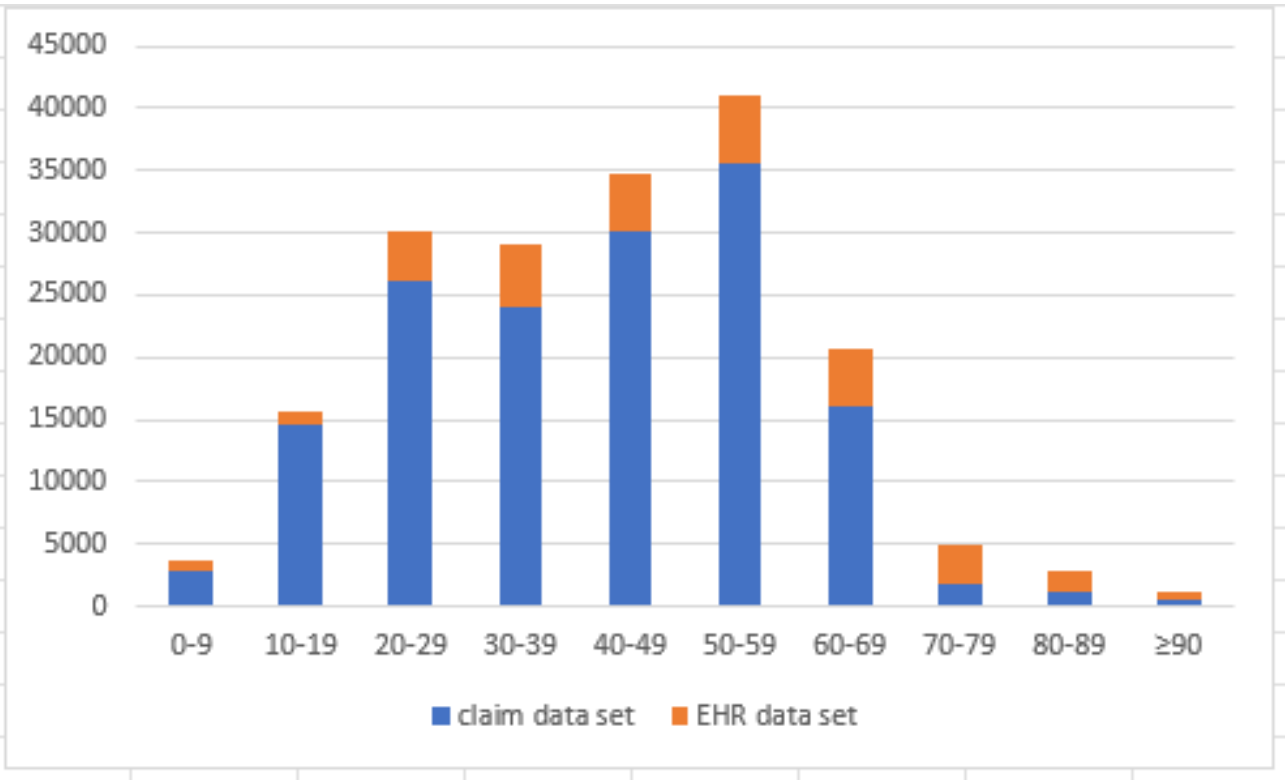
Descriptive statistics are summarized as frequencies and percentages for categorical data. A simple (or unadjusted) logistic regression model was used to assess the strength of the association between demographic and clinical factors and phenotype. The demographic and clinical factors included demographics, comorbidities, and interventions. In this study, phenotype was defined as a dichotomous variable, and we focused on diagnosis of a thrombotic complication (ie, with or without thrombotic complication).

The results are summarized by the odds ratio (OR), corresponding 95% CI, and *P* value. All tests were 2-sided and the significance level was set to $P=.001$. All statistical analyses were performed using the Modern Applied Statistics with S (MASS) statistical software library version 7.3.54 [29] in R, version 4.1.0 [30].

Age and Gender Distributions From Patients in the Claims and EHR Data Sets

Within the study time period, there were 153,137 COVID-19 patients in the claim data, with 44.8% being men. There were 31,364 COVID-19 patients in the EHR data, with 43.9% being men. The age distributions are shown in Figure 1.

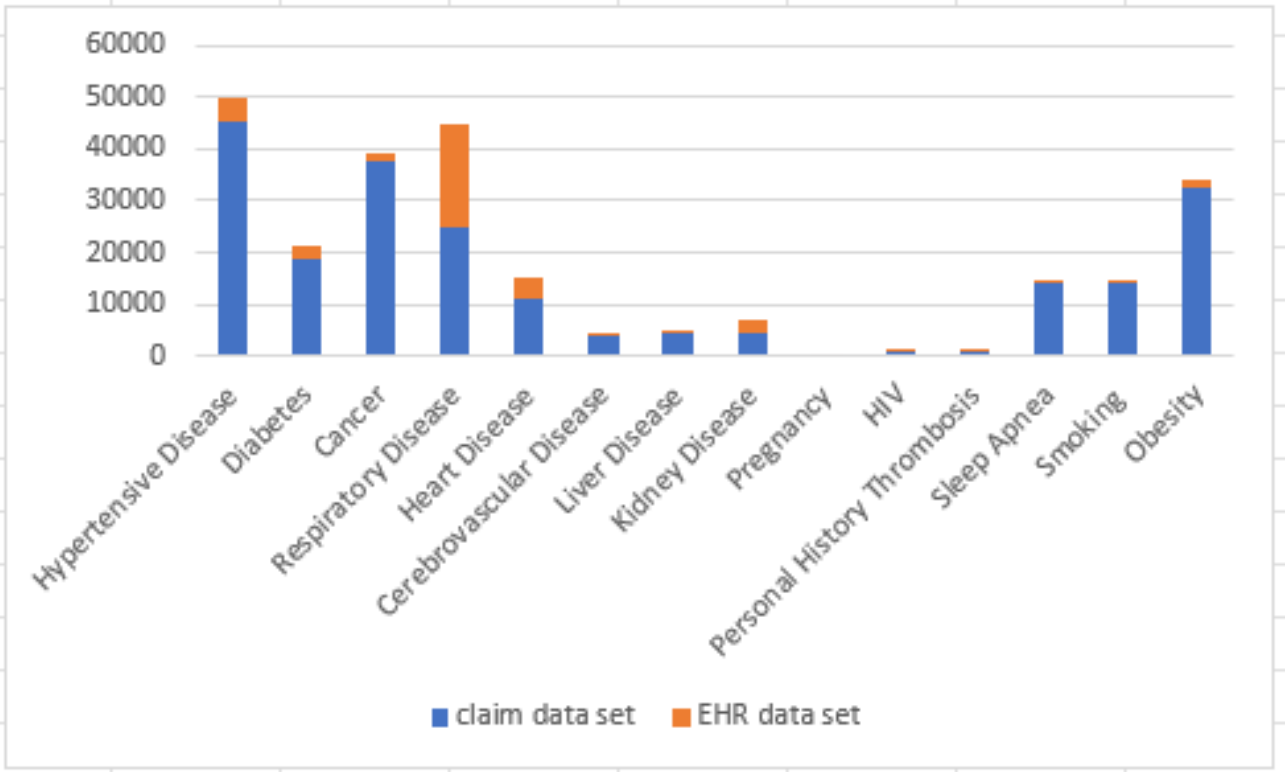
Figure 1. Patients' age distributions from the insurance claims and electronic health record (EHR) data sets. The x-axis is age and the y-axis is patient count.



Comorbidity Distributions From Patients in the Claims and EHR Data Sets

COVID-19 patient comorbidity distributions are shown in Figure 2.

Figure 2. Patients' comorbidity distributions from the insurance claims and electronic health record (EHR) data sets. The x-axis is comorbidities and the y-axis is patient counts.



Handling of Missing Data

We encountered two types of missing data. The first involved missing at least one variable in a data set. Given the low rate of such missingness (<2.5% for any individual variable), imputation was deemed unnecessary [28,31]. The other involved missing a category of data in one data set. There were three such cases: ethnicity and lab data are in the EHR data set but not in the claims dataset, whereas region data are in the claims data set but not in the EHR data set (the EHR data set includes patients mostly from the northeast region of the United States). We performed analysis on one data set in the three cases with the understanding that they would not be cross-verified.

Ethics Approval

This study was approved by the institutional review board of MGB (IRB Protocol #2021P001133).

Data Analysis

We performed an analysis to determine patients' clinical and demographic factors associated with thrombotic complications for patients with a COVID-19 diagnosis. Data queried from IBM MarketScan were stored as CSV files. The analysis read from the CSV files and drew hypotheses based on a predefined *P* value threshold (<.001), which was then verified using data queried from the RPDR database.

Results

Age and Thrombotic Complications

To compare the thrombotic complications between the young and old population, we categorized COVID-19 patients into two age groups: those younger than 60 years and those aged 60 years and older. Table 1 lists the frequency (ie, count) of COVID-19 patients with and without thrombotic complications, the calculated *P* value, OR, and the 95% CI from the claims database. The corresponding data from the EHR-compatible database are listed in Table 2. As demonstrated, age and thrombotic complications were significantly associated. In addition, patients aged 60 years and older had a much higher odds to have thrombotic complications. Results from both data sets were consistent despite patients in the two data sets being from different geographical regions and backgrounds. This provided more confidence to the findings and showed how the two data sets could enrich each other.

As shown in Figure 3, with finer age grouping, we also observed that the OR for thrombotic complications consistently increased with age (with the exception that the odds for the age groups of 80-89 years and 90 years and older were similar), with *P*<.001.

Table 1. Age and thrombotic complications and the strength of their association based on claims data.

Age group	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
<60 years	130,293	3314	Reference	N/A[a]
≥60 years	17,379	2151	4.866 (4.599-5.149)	<.001

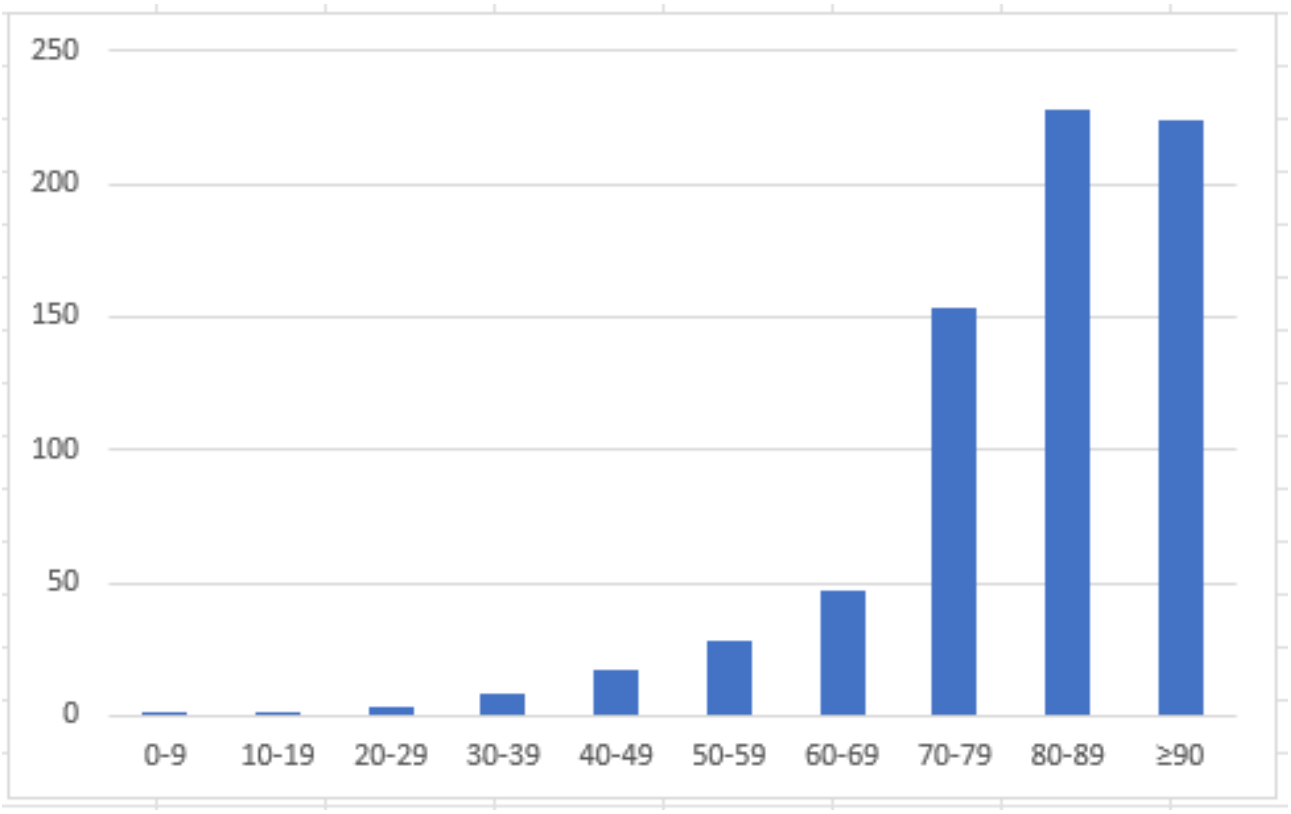
[a]N/A: not applicable.

Table 2. Age and thrombotic complications and the strength of their association based on electronic health record data.

Age group	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
<60 years	20,338	487	Reference	N/A[a]
≥60 years	8796	1339	6.357 (5.714-7.073)	<.001

[a]N/A: not applicable.

Figure 3. Odds ratio of thrombotic complications with age (*P*<.001). The x-axis is age and the y-axis is the odds ratio.



Gender and Thrombotic Complications

Men had higher odds for thrombotic complications when compared to women from both data sets (Tables 3 and 4). Similar to age, the results showed that the two data sets were consistent and enrich each other. This result aligns with prior literature [32].

Table 3. Gender and thrombotic complications and the strength of their association based on claims data.

Gender	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Men	65,926	2738	1.245 (1.180-1.314)	<.001
Women	81,188	2727	Reference	N/A[a]

[a]N/A: not applicable.

Table 4. Gender and thrombotic complications and the strength of their association based on electronic health record data.

Gender	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Men	12,708	1064	1.693 (1.542-1.859)	<.001
Women	16,763	829	Reference	N/A[a]

[a]N/A: not applicable.

Comorbidities and Thrombotic Complications

We examined the associations between a thrombotic complication and preexisting conditions such as hypertensive disease, diabetes, cancer, respiratory disease, heart disease, cerebrovascular disease, liver disease, pregnancy, HIV, personal history of thrombosis, sleep apnea, smoking, and obesity. All comorbidities were significantly associated with thrombosis in both data sets (Tables 5 and 6). Although the relative ORs differed in the two data sets, the results were consistent. In both data sets, patients with cerebrovascular disease had the second highest odds to have thrombotic complications, and patients with heart diseases had very similar odds to have a thrombotic complication. In addition, patients with HIV, cancer, and obesity had relatively lower odds to have thrombosis in both data sets. A major difference was that personal history of thrombosis had the highest odds in the claims data set but ranked fifth in the EHR data set. This difference might be due to the small number of patients (n=250) with a personal history of thrombosis in the EHR data set.

Table 5. Comorbidity and thrombotic complications and the strength of their association based on claims data.

Comorbidity	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Hypertension	41,513	3890	6.316 (6.950-6.704)	<.001
Diabetes	16,688	1959	4.386 (4.140-4.646)	<.001
Cancer	35,684	2092	1.947 (1.841-2.058)	<.001
Respiratory disease	23,156	1847	2.745 (2.591-2.908)	<.001
Heart disease	8743	2401	12.452 (11.755-13.191)	<.001
Cerebrovascular disease	2542	1406	19.776 (18.399-21.258)	<.001
Liver disease	4044	450	3.187 (2.880-3.527)	<.001
Kidney disease	3316	989	9.619 (8.906-10.389)	<.001
HIV	712	51	1.944 (1.462-2.587)	<.001
History of thrombosis	189	473	73.938 (62.318-87.727)	<.001
Sleep apnea	12,970	1178	2.854 (2.669-3.051)	<.001
Smoking use	13,141	1177	2.810 (2.628-3.005)	<.001
Obesity	30,288	2293	2.802 (2.651-2.961)	<.001

Table 6. Comorbidity and thrombotic complications and the strength of their association based on electronic health record data.

Comorbidity	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Hypertension	3079	1172	13.675 (12.373-15.113)	<.001
Diabetes	1894	671	7.853 (7.070-8.723)	<.001
Cancer	890	261	5.048 (4.359-5.845)	<.001
Respiratory disease	17,858	1717	5.891 (5.045-6.879)	<.001
Heart disease	2768	1118	13.661 (12.366-15.093)	<.001
Cerebrovascular disease	294	253	15.053 (12.632-17.937)	<.001
Liver disease	416	160	6.340 (5.25-7.656)	<.001
Kidney disease	2144	718	7.648 (6.902-8.476)	<.001
HIV	60	17	4.368 (2.544-7.499)	<.001
History of thrombosis	159	91	9.154 (7.044-11.896)	<.001
Sleep apnea	395	155	6.454 (5.327-7.820)	<.001
Smoking use	159	223	24.206 (19.633-29.842)	<.001
Obesity	1073	237	3.722 (3.207-4.321)	<.001

External Intervention and Thrombotic Complications

We examined three external interventions (veno-venous ECMO, mechanical ventilation, and extraneous oxygen use) and their association with thrombotic complication. The ORs and *P* values are summarized in Table 7 and Table 8 for claims and EHR compatible data sets, respectively. Veno-venous ECMO and extraneous oxygen interventions were strongly associated with thrombotic complications in both data sets. Mechanical ventilation was significantly associated with thrombotic complications in the claims data set; however, the number of cases in the EHR compatible data set was too low for appropriate analysis.

Table 7. External interventions and thrombotic complications and the strength of their association based on claims data.

External Interventions	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Veno-venous ECMO[a]	34	36	28.794 (18.005-46.047)	<.001
Mechanical ventilation	372	324	24.955 (21.447-29.037)	<.001
Extraneous oxygen use	423	231	15.364 (13.057-18.078)	<.001

[a]ECMO: extracorporeal membrane oxygenation.

Table 8. External interventions and thrombotic complications and the strength of their association based on electronic health record data.

External interventions	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Veno-venous ECMO[a]	28	25	13.839 (8.054-23.779)	<.001
Mechanical ventilation	3	3	15.332 (3.092-76.016)	<.001
Extraneous oxygen use	137	56	6.418 (4.687-8.790)	<.001

[a]ECMO: extracorporeal membrane oxygenation.

Medication Intervention and Thrombotic Complications

We examined six medication interventions (lopinavir/ritonavir, dexamethasone, remdesivir, monoclonal antibody, tocilizumab, and antimalarials) and their association with thrombotic complication using EHR data. The ORs and *P* values are summarized in Table 9 for the EHR data set. Approximately 1.74% of the COVID-19 patients, the highest proportion in this group, took dexamethasone. A previous report showed that dexamethasone was associated with a reduction in mortality in patients with advanced COVID-19 [33]. Our analysis showed that these patients are 5 times more likely to have thrombotic complications. Approximately 1.26% of the COVID-19 patients, the second highest proportion in this group, took remdesivir. Remdesivir was suggested to be beneficial in shortening the time to recovery in hospitalized COVID-19 patients [34]. Our analysis showed that these patients are also 3 times more likely to have thrombotic complications.

For the claims data set, information was available for three of the above medicines, and the results of this analysis are shown in Table 10. Approximately 2.71% of the COVID-19 patients, the highest proportion in this group, took dexamethasone, and they were 3 times more likely to have thrombotic complications.

Table 9. Medication and thrombotic complications and the strength of their association based on electronic health record data.

Medication interventions	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Lopinavir/ritonavir	10	3	4.599 (1.265-16.723)	.02
Dexamethasone	405	134	5.375 (4.396-6.573)	<.001
Remdesivir	325	66	3.185 (2.434-4.168)	<.001
Monoclonal antibody	18	15	12.852 (6.467-25.541)	<.001
Tocilizumab	119	37	4.835 (3.333-7.013)	<.001
Antimalarials	3	3	15.332 (3.093-76.016)	.001

Table 10. Medication and thrombotic complications and the strength of their association based on claims data.

Medication interventions	No Thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Lopinavir/ritonavir	5	3	16.221 (3.876-67.893)	<.001
Dexamethasone	3706	442	3.418 (2.083-3.788)	<.001
Antimalarials	2346	177	2.074 (1.775-2.422)	<.001

Lab Results and Thrombotic Complications

We examined six lab results that were recorded as abnormal from the EHR data set. The results of the analysis are summarized in Table 11. The claims data set does not have corresponding lab information.

Table 11. Strength of associations between lab results and thrombotic complications based on electronic health record data.

Lab result	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
D-dimer level	5354	1418	13.174 (11.824-14.677)	<.001
Platelet count	14,279	1807	21.634 (17.404-26.891)	<.001
Prothrombin time	5635	1528	17.344 (15.416-19.513)	<.001
Fibrin degradation products	5544	1208	7.455 (6.758-8.224)	<.001
Fibrinogen	4183	1117	8.533 (7.742-9.405)	<.001
C-reactive protein	12,439	1688	10.95 (9.455-12.682)	<.001

Ethnicity and Thrombotic Complications

We examined ethnicities and their associations with thrombotic complications using the EHR data set (Table 12). The claims data set does not have ethnicity-related information.

Table 12. Strength of associations between ethnicity and thrombotic complications based on electronic health record data.

Ethnicities	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Asian	1007	53	0.800 (0.575-1.012)	.06
Black	3293	285	1.357 (1.181-1.558)	<.001
Hispanic	1770	76	0.606 (0.478-0.768)	<.001
White	17,527	1193	Reference	N/A[a]

[a]N/A: not applicable.

Region and Thrombotic Complications

The insurance claims data set includes patients from all regions. With *P*<.001, the Northcentral region had the highest OR and the West had the lowest OR for thrombotic complications (Table 13). The EHR data set includes mostly patients in the Northeast where the MGB is located, and therefore the corresponding analysis was not available.

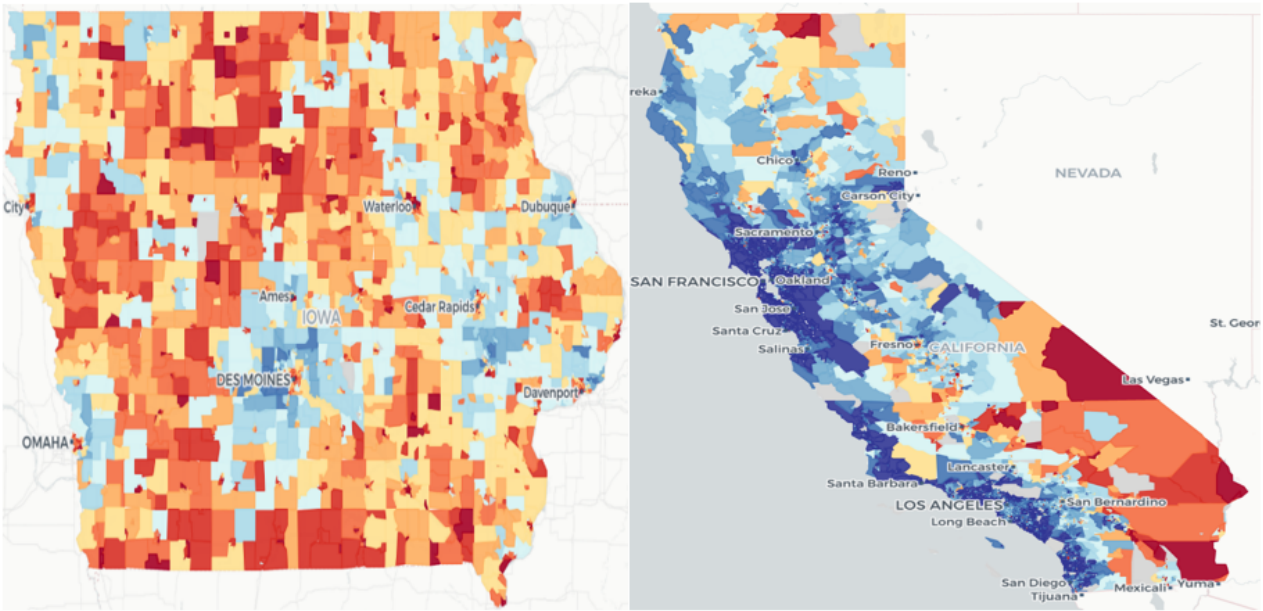
The regions are divided as shown in Multimedia Appendix 1. Our analysis demonstrated that COVID-19 patients in the Northcentral region of the United States have an OR of 1.562, while patients in the West have an OR of 0.701 to have thrombotic complications. This finding correlates well with the Area Deprivation Index (ADI) of regions [35]. We have included the ADI values of Iowa (Northcentral) and California (West) in Figure 4 to highlight this point.

Table 13. Strength of associations between region and thrombotic complications based on insurance claims data.

Region	No thrombotic complication, n	Thrombotic complication, n	Odds ratio (95% CI)	*P* value
Northeast	27,742	1007	Reference	N/A[a]
Northcentral	25,912	1631	1.562 (1.439-1.695)	<.001
South	78,557	2588	0.908 (0.843-0.977)	<.001
West	14,964	381	0.701 (0.622-0.791)	<.001

[a]N/A: not applicable.

Figure 4. Area Deprivation Index of Iowa (left) and California (right). Deep red indicates the most disadvantaged area and deep blue indicates the least disadvantaged area. Iowa belongs to the Northcentral region where COVID-19 patients have an odds ratio of 1.562 of having thrombotic complications. California belongs to the West region, where COVID-19 patients have an odds ratio of 0.701 of having thrombotic complications.



Discussion

Principal Findings

We found factors related to the demographics, comorbidities, therapeutic interventions, and labs of COVID-19 patients that are strongly associated with the risk of experiencing thrombotic complications. The analysis approach adopted in this study can be leveraged to work across heterogeneous patient databases from different health care and research organizations by using deidentified patient count data. This study used claims and EHR data sets as a case study, but the approach can also be generalized to handle multiple data sources.

The counts were queried with ICD-10 diagnosis codes of the phenotypes being studied. This facilitates collaboration in tackling difficult local and global health issues. In this case study, we analyzed thrombotic complications associated with demographic and clinical factors in COVID-19 patients using insurance claims and EHR databases. We found the design to be very productive in our collaboration where we used claims data to draw hypotheses and EHR data for validation. The two data sets are mostly consistent and enrich each other, except in cases for a very small sample size in EHR-derived data.

The claims and EHR databases have different storage formats, query syntaxes, and security concerns. Our design was to use the common ICD-10 code to run queries on each database and store the query results in CSV files, so that we could use the same R code to read the CSV files and perform the statistical analysis. This also minimized data exchanges between the two geographically dispersed teams.

When selecting factors that are associated with thrombotic complications, we focused on four main categories: demographics, comorbidities, interventions, and lab results. A problem we encountered was that some categories of data might be missing from one data set; for example, the claims database does not have information on all the same prescription drugs as found in the EHR data set. We performed the analysis using only one data set when we deemed the factors of interest to be potentially important. Our analysis of the EHR data set showed that the three most frequently used medications are dexamethasone, remdesivir, and tocilizumab. These medications were associated with thrombotic complications with ORs of 5.375, 3.185, and 4.835, respectively. These were also the medications considered in a previous model developed to predict the requirement of ICU and VTE for COVID-19 patients [28]. Patient lab data, including the D-dimer level, platelet count, prothrombin time, fibrin degradation products, and fibrinogen, are available only in the EHR data set. We believed that these factors are clinically associated with thrombotic complications, which was supported by our analysis results. D-dimer level, one of the top three factors in the lab results category, had an OR of 13 for thrombotic complications, and was also used to predict VTE development in COVID-19 patients in the previous study [28]. All of the top three findings, D-dimer level, platelet count, and prothrombin time, were also previously used in a machine-learning model to predict the need for invasive mechanical ventilation and the mortality of COVID-19 patients [36]. This further validated the strength of the model when applied to large and diverse data sets.

All patients with the preexisting conditions listed in Tables 5 and 6 had much higher odds of having thrombotic complications than other patients in both data sets. It is interesting to note that for patients with underlying cerebrovascular disease, the odds of thrombotic complications in insurance claims data were 19-fold higher and the odds in EHR-derived data were 15-fold higher, ranking second in the comorbidities for COVID-19 patients. It is also interesting to note that for patients with heart disease, the odds for thrombotic complications in both data sets were approximately 13-fold higher, and both ranked third in the comorbidity listings. This further highlights the consistency of the two data sets.

For COVID-19 patients that received external interventions of veno-venous ECMO and extraneous oxygen use, each intervention had much higher odds for thrombotic complications in both data sets.

MarketScan claims data can potentially be very useful in understanding the impact of COVID-19 by monitoring these cases longitudinally to document short-term and long-term patient outcomes.

Comparison to Prior Work

We found that COVID-19 patients aged 60 years and older were approximately 5 times more likely to have thrombotic complications than those under 60 years old. Although it is well documented that older patients are more susceptible to thrombotic complications [37], this research provides a quantitative measurement of the degree to which this is true in COVID-19 patients.

In terms of gender, men were 1.25 times more likely in the claims data (and 1.69 times more likely in EHR-derived data) to have thrombotic complications compared to women. Although the ORs were slightly different between the two data sets, both showed that men are statistically more likely than women to have thrombotic complications. This finding is consistent with previous studies indicating that men are more likely to be afflicted with thrombotic complications [32].

Strengths and Limitations

This study used two distinct data sets with 184,831 COVID-19 patients and very comprehensive demographic and clinical information. This allowed us to investigate thrombotic complications from different aspects. We designed an approach that worked with both data sets and found factors strongly associated with thrombotic complications. This approach facilitated teams with different data formats to collaborate. Furthermore, our findings are consistent with the existing literature.

This study focused on patients who received a COVID-19 diagnosis between February 1, 2020, and September 30, 2020, in the United States, and the EHR-derived data included mostly patients in the northeast region of the country. Thus, this data source does not cover the full domestic United States or global perspective. Although we used data from over 184,000 COVID-19 patients and a very small P value threshold (P<.001)

to draw and verify hypotheses on whether a clinical factor affected thrombotic complications, the overall patient count used is relatively small compared with the global patient counts.

We examined factors individually, but it is possible that some factors might be correlated. This was the first phase of the research, and the main goal was to verify the consistency of the two data sets, demonstrating that all factors are associated with thrombotic complications. The second phase of this research will focus on multivariables analysis, as described below. Moreover, this study did not investigate the temporal relationship between interventions and the thrombosis complications.

Future Directions

To determine how each factor contributes to a patient's thrombotic complications, we will explore explainable machine-learning models [38-40] to train models with all the factors we identified in this study. The databases can provide deidentified individual patient data, which can be used to train explainable machine-learning models. The models will not only predict a COVID-19 patient's risk of thrombotic complications but also determine each factor's contribution.

Conclusions

In this work, we examined heterogeneous patient databases and performed an analysis that does not depend on individual patient–level data. This proved to be a valuable approach for collaboration between health care and research organizations with data from different sources, in different storage formats, and with different patient privacy constraints. Via analysis across research collaborators with heterogeneous data sources, we found important demographic and clinical factors associated with thrombotic complications in patients with COVID-19. Our research provides for a collaborative and early risk stratification approach, as a critical step toward helping to ensure efficient resource allocation and better outcomes for the COVID-19 patient population.

Authors' Contributions

All authors collaborated in designing the study, collecting data, and performing statistical analysis and results explanation. AZ drafted the initial manuscript, and all authors provided critical comments, revised, and approved the manuscript in its final form for submission.

Conflicts of Interest

BR is an IBM employee.

Multimedia Appendix 1
Division of US regions.
[DOCX File , 12 KB - jmir_v24i10e35860_app1.docx]

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. 2021. URL: https://covid19.who.int [accessed 2022-09-02]
2. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
3. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical predictive models for COVID-19: systematic study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]
4. Makridis CA, Strebel T, Marconi V, Alterovitz G. Designing COVID-19 mortality predictions to advance clinical outcomes: evidence from the Department of Veterans Affairs. BMJ Health Care Inform 2021 Jun 09;28(1):e100312 [FREE Full text] [doi: 10.1136/bmjhci-2020-100312] [Medline: 34108143]
5. Hao B, Sotudian S, Wang T, Xu T, Hu Y, Gaitanidis A, et al. Early prediction of level-of-care requirements in patients with COVID-19. Elife 2020 Oct 12;9:e60519. [doi: 10.7554/eLife.60519] [Medline: 33044170]
6. Wollenstein-Betech S, Cassandras C, Paschalidis I. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: hospitalizations, mortality, and the need for an ICU or ventilator. medRxiv. 2020 May 08. URL: https://www.medrxiv.org/content/10.1101/2020.05.03.20089813v1 [accessed 2022-09-02]
7. Kasturi SN, Park J, Wild D, Khan B, Haggstrom DA, Grannis S. Predicting COVID-19-related health care resource utilization across a statewide patient population: model development study. J Med Internet Res 2021 Nov 15;23(11):e31337 [FREE Full text] [doi: 10.2196/31337] [Medline: 34581671]
8. Pizzolo F, Rigoni AM, De Marchi S, Friso S, Tinazzi E, Sartori G, et al. Deep vein thrombosis in SARS-CoV-2 pneumonia-affected patients within standard care units: Exploring a submerged portion of the iceberg. Thromb Res 2020 Oct;194:216-219 [FREE Full text] [doi: 10.1016/j.thromres.2020.08.008] [Medline: 33074107]
9. Avila J, Long B, Holladay D, Gottlieb M. Thrombotic complications of COVID-19. Am J Emerg Med 2021 Jan;39:213-218 [FREE Full text] [doi: 10.1016/j.ajem.2020.09.065] [Medline: 33036855]

10. Khan IH, Savarimuthu S, Leung MST, Harky A. The need to manage the risk of thromboembolism in COVID-19 patients. J Vasc Surg 2020 Sep;72(3):799-804 [FREE Full text] [doi: 10.1016/j.jvs.2020.05.015] [Medline: 32417304]
11. Ronderos Botero DM, Omar AMS, Sun HK, Mantri N, Fortuzi K, Choi Y, et al. COVID-19 in the healthy patient population: demographic and clinical phenotypic characterization and predictors of in-hospital outcomes. Arterioscler Thromb Vasc Biol 2020 Nov;40(11):2764-2775 [FREE Full text] [doi: 10.1161/ATVBAHA.120.314845] [Medline: 32907371]
12. Arnardottir H, Pawelzik SC, Sarajlic P, Quaranta A, Kolmert J, Religa D, et al. Immunomodulation by intravenous omega-3 fatty acid treatment in older subjects hospitalized for COVID-19: a single-blind randomized controlled trial. MedRxiv. URL: https://www.medrxiv.org/content/10.1101/2021.12.27.21268264v1 [accessed 2022-09-02]
13. Bikdeli B, Madhavan MV, Jimenez D, Chuich T, Dreyfus I, Driggin E, Global COVID-19 Thrombosis Collaborative Group, Endorsed by the ISTH, NATF, ESVM,the IUA, Supported by the ESC Working Group on Pulmonary CirculationRight Ventricular Function. COVID-19 and thrombotic or thromboembolic disease: implications for prevention, antithrombotic therapy, and follow-up: JACC state-of-the-art review. J Am Coll Cardiol 2020 Jun 16;75(23):2950-2973 [FREE Full text] [doi: 10.1016/j.jacc.2020.04.031] [Medline: 32311448]
14. Klok F, Kruip M, van der Meer N, Arbous M, Gommers D, Kant K, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res 2020 Jul;191:145-147 [FREE Full text] [doi: 10.1016/j.thromres.2020.04.013] [Medline: 32291094]
15. Cui S, Chen S, Li X, Liu S, Wang F. Prevalence of venous thromboembolism in patients with severe novel coronavirus pneumonia. J Thromb Haemost 2020 Jun 06;18(6):1421-1424. [doi: 10.1111/jth.14830] [Medline: 32271988]
16. Pellicori P, Doolub G, Wong CM, Lee KS, Mangion K, Ahmad M, et al. COVID-19 and its cardiovascular effects: a systematic review of prevalence studies. Cochrane Database Syst Rev 2021 Mar 11;3:CD013879 [FREE Full text] [doi: 10.1002/14651858.CD013879] [Medline: 33704775]
17. Kollias A, Kyriakoulis KG, Lagou S, Kontopantelis E, Stergiou GS, Syrigos K. Venous thromboembolism in COVID-19: A systematic review and meta-analysis. Vasc Med 2021 Aug;26(4):415-425 [FREE Full text] [doi: 10.1177/1358863X21995566] [Medline: 33818197]
18. Minet C, Potton L, Bonadona A, Hamidfar-Roy R, Somohano CA, Lugosi M, et al. Venous thromboembolism in the ICU: main characteristics, diagnosis and thromboprophylaxis. Crit Care 2015 Aug 18;19(1):287 [FREE Full text] [doi: 10.1186/s13054-015-1003-9] [Medline: 26283414]
19. Malato A, Dentali F, Siragusa S, Fabbiano F, Kagoma Y, Boddi M, et al. The impact of deep vein thrombosis in critically ill patients: a meta-analysis of major clinical outcomes. Blood Transfus 2015 Oct;13(4):559-568. [doi: 10.2450/2015.0277-14] [Medline: 26513770]
20. Labenz C, Kremer WM, Schattenberg JM, Wörns MA, Toenges G, Weinmann A, et al. Clinical Frailty Scale for risk stratification in patients with SARS-CoV-2 infection. J Investig Med 2020 Aug 07;68(6):1199-1202 [FREE Full text] [doi: 10.1136/jim-2020-001410] [Medline: 32641351]
21. Zöller B, Li X, Sundquist J, Sundquist K. Autoimmune diseases and venous thromboembolism: a review of the literature. Am J Cardiovasc Dis 2012;2(3):171-183 [FREE Full text] [Medline: 22937487]
22. Yina W. Application of EHR in health care. 2010 Presented at: 2010 Second International Conference on Multimedia and Information Technology; April 24-25, 2010; Washington, DC. [doi: 10.1109/mmit.2010.32]
23. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med 2021 Apr;27(4):601-615 [FREE Full text] [doi: 10.1038/s41591-021-01283-z] [Medline: 33753937]
24. Bilaloglu S, Aphinyanaphongs Y, Jones S, Iturrate E, Hochman J, Berger JS. Thrombosis in hospitalized patients With COVID-19 in a New York City health system. JAMA 2020 Aug 25;324(8):799-801 [FREE Full text] [doi: 10.1001/jama.2020.13372] [Medline: 32702090]
25. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language processing. J Med Internet Res 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: 10.2196/21801] [Medline: 33090964]
26. Vollmer Dahlke D, 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 Mar 27;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
27. INSPIRATION Investigators, Sadeghipour P, Talasaz AH, Rashidi F, Sharif-Kashani B, Beigmohammadi MT, et al. Effect of intermediate-dose vs standard-dose prophylactic anticoagulation on thrombotic events, extracorporeal membrane oxygenation treatment, or mortality among patients with COVID-19 admitted to the intensive care unit: The INSPIRATION randomized clinical trial. JAMA 2021 Apr 27;325(16):1620-1630 [FREE Full text] [doi: 10.1001/jama.2021.4152] [Medline: 33734299]
28. Shah S, Switzer S, Shippee ND, Wogensen P, Kosednar K, Jones E, et al. Implementation of an anticoagulation practice guideline for COVID-19 via a clinical decision support system in a large academic health system and its evaluation: observational study. JMIR Med Inform 2021 Nov 18;9(11):e30743 [FREE Full text] [doi: 10.2196/30743] [Medline: 34550900]
29. Venables WN, Ripley BD. Modern applied statistics with S. Fourth edition. New York, NY: Springer; 2002.

30. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL: http://www.R-project.org [accessed 2022-09-06]
31. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
32. Bauersachs RM, Riess H, Hach-Wunderle V, Gerlach H, Carnarius H, Eberle S, et al. Impact of gender on the clinical presentation and diagnosis of deep-vein thrombosis. Thromb Haemost 2010 Apr;103(4):710-717. [doi: 10.1160/TH09-10-0705] [Medline: 20174769]
33. Jensen MP, George M, Gilroy D, Sofat R. Beyond dexamethasone, emerging immuno-thrombotic therapies for COVID-19. Br J Clin Pharmacol 2021 Mar 14;87(3):845-857. [doi: 10.1111/bcp.14540] [Medline: 32881064]
34. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, ACTT-1 Study Group Members. Remdesivir for the Treatment of Covid-19 - Final Report. N Engl J Med 2020 Nov 05;383(19):1813-1826 [FREE Full text] [doi: 10.1056/NEJMoa2007764] [Medline: 32445440]
35. Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med 2014 Dec 02;161(11):765-774 [FREE Full text] [doi: 10.7326/M13-2946] [Medline: 25437404]
36. Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, et al. COVID-19 mortality prediction from deep learning in a large multistate electronic health record and laboratory information system data set: algorithm development and validation. J Med Internet Res 2021 Sep 28;23(9):e30157 [FREE Full text] [doi: 10.2196/30157] [Medline: 34449401]
37. Engbers MJ, van Hylckama Vlieg A, Rosendaal FR. Venous thrombosis in the elderly: incidence, risk factors and risk groups. J Thromb Haemost 2010 Oct;8(10):2105-2112. [doi: 10.1111/j.1538-7836.2010.03986.x] [Medline: 20629943]
38. Ploug T, Sundby A, Moeslund TB, Holm S. Population preferences for performance and explainability of artificial intelligence in health care: choice-based conjoint survey. J Med Internet Res 2021 Dec 13;23(12):e26611 [FREE Full text] [doi: 10.2196/26611] [Medline: 34898454]
39. Ammar N, Shaban-Nejad A. Explainable artificial intelligence recommendation system by leveraging the semantics of adverse childhood experiences: proof-of-concept prototype development. JMIR Med Inform 2020 Nov 04;8(11):e18752 [FREE Full text] [doi: 10.2196/18752] [Medline: 33146623]
40. Zhang A, Teng L, Alterovitz G. An explainable machine learning platform for pyrazinamide resistance prediction and genetic feature identification of Mycobacterium tuberculosis. J Am Med Inform Assoc 2021 Mar 01;28(3):533-540 [FREE Full text] [doi: 10.1093/jamia/ocaa233] [Medline: 33215194]

Abbreviations

ADI: Area Deprivation Index
COBRA: Consolidated Omnibus Budget Reconciliation Act
ECMO: extracorporeal membrane oxygenation
EHR: electronic health record
ICD-10: International Classification of Diseases, Tenth Revision
ICU: intensive care unit
MASS: Modern Applied Statistics with S
MGB: Mass General Brigham
OR: odds ratio
RPDR: Research Patient Data Registry
STEMI: ST elevation myocardial infarction
VTE: venous thromboembolism

Edited by T Leung; submitted 21.12.21; peer-reviewed by P Sarajlic, K Fultz Hollis, Y Cao; comments to author 17.02.22; revised version received 06.05.22; accepted 17.05.22; published 21.10.22.

Please cite as:
Rosario B, Zhang A, Patel M, Rajmane A, Xie N, Weeraratne D, Alterovitz G
Characterizing Thrombotic Complication Risk Factors Associated With COVID-19 via Heterogeneous Patient Data: Retrospective Observational Study
J Med Internet Res 2022;24(10):e35860
URL: https://www.jmir.org/2022/10/e35860
doi:10.2196/35860
PMID:36044652

Original Paper

Physical Therapists' Knowledge and Attitudes Regarding Artificial Intelligence Applications in Health Care and Rehabilitation: Cross-sectional Study

Mashael Alsobhi[1], PhD; Fayaz Khan[1], PhD; Mohamed Faisal Chevidikunnan[1], PhD; Reem Basuodan[2], PhD; Lama Shawli[3], BPT; Ziyad Neamatallah[1], PhD

[1]Department of Physical Therapy, Faculty of Medical Rehabilitation Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

[2]Department of Rehabilitation Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[3]Department of Occupational Therapy, College of Applied Medical Sciences, King Saud bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia

Corresponding Author:
Fayaz Khan, PhD
Department of Physical Therapy
Faculty of Medical Rehabilitation Sciences
King Abdulaziz University
Administration Street
Jeddah, 22254
Saudi Arabia
Phone: 966 533034058
Email: fayazrkhan@gmail.com

Abstract

Background: The use of artificial intelligence (AI) in the field of rehabilitation is growing rapidly. Therefore, there is a need to understand how physical therapists (PTs) perceive AI technologies in clinical practice.

Objective: This study aimed to investigate the knowledge and attitude of PTs regarding AI applications in rehabilitation based on multiple explanatory factors.

Methods: A web-based Google Form survey, which was divided into 4 sections, was used to collect the data. A total of 317 PTs participated voluntarily in the study.

Results: The PTs' knowledge about AI applications in rehabilitation was lower than their knowledge about AI in general. We found a statistically significant difference in the PTs' knowledge regarding AI applications in the rehabilitation field based on sex (odds ratio [OR] 2.43, 95% CI 1.53-3.87; $P<.001$). In addition, experience (OR 1.79, 95% CI 1.11-2.87; $P=.02$) and educational qualification (OR 1.68, 95% CI 1.05-2.70; $P=.03$) were found to be significant predictors of knowledge about AI applications. PTs who work in the nonacademic sector and who had <10 years of experience had positive attitudes regarding AI.

Conclusions: AI technologies have been integrated into many physical therapy practices through the automation of clinical tasks. Therefore, PTs are encouraged to take advantage of the widespread development of AI technologies and enrich their knowledge about, and enhance their practice with, AI applications.

(J Med Internet Res 2022;24(10):e39565) doi:10.2196/39565

KEYWORDS

artificial intelligence; physical therapy; clinicians' attitudes; health care; rehabilitation; digital health; machine learning; survey

Introduction

Background

The use of artificial intelligence (AI) has been growing rapidly in the fields of health care and rehabilitation [1]. Many of AI's clinical benefits have been mentioned in the literature. AI is defined as the ability of a machine to perform a functional task moderated intelligently by humans [2]. AI uses algorithms to learn, think, and then assist in various clinical practices such as radiology [3], dentistry [4], dermatology [5], and rehabilitation [6]. In addition, AI provides up-to-date clinical information

from scientific resources such as journals, books, and evidence-based practice, which assists health care providers in clinical decision-making. Furthermore, AI technologies help to reduce medical errors in daily human practices [7-9].

Today, AI technologies are used in multidisciplinary health care research fields, and researchers are exploring and investigating the practical implications of using such technologies. In rehabilitation, AI has been used to enhance the patient care process by assisting physical therapists (PTs) either in providing a comprehensive assessment or in predicting patients' performance or determining a diagnosis [10]. Moreover, research has revealed more uses of AI in medical and rehabilitation practices, such as problem solving, x-ray diagnosis, planning treatment protocols, and physical manipulation of patients [11]. All these functions of AI are core elements of physical therapy professional practice. Consequently, it is worth stating that many physical therapy practices might be susceptible to automation by AI technologies. In the study by Brougham and Haar [12], futurists are quoted as predicting that a third of the jobs that exist today could be taken by smart technology, artificial intelligence, robotics, and algorithms by 2025.

Machine learning (ML), a subset of AI, enables practitioners to use known quantities from data to make predictions [13]. In addition, ML is used to enable computerized decision-making and provide predictions based on patient data, and it can also be used as a tool to provide immediate preventive care for patients with specific conditions [14]. In 2020, a study was conducted by Ye et al [15] to validate a tool that was developed based on ML algorithms to predict older adults' fall risk. The researchers found that the ML-based fall risk tool was a valid tool for producing automatic early warnings, which may prevent falls among older adults. In fact, patients with orthopedic and neurological disorders need an intensive rehabilitation physical therapy program that might last for months to improve their functional disabilities. Subsequently, PTs might face challenges in designing therapeutic interventions based on their understanding of the patients' performance. In such cases, an ML-based AI decision support system would help PTs in determining diagnosis and monitoring the rehabilitation intervention.

By contrast, as AI technologies become more widespread, the need for AI education among PTs becomes essential. A qualitative study was conducted in the United Kingdom by Castagno and Khalifa in 2020 [2] to explore the knowledge and attitudes of health care providers regarding current and future uses of AI. The researchers reported a lack of full understanding of AI fundamentals as well as concerns about the potential consequences of the use of AI in clinical practice among health care professionals. Given the fact that AI technologies may perform some of the PTs' work, it is necessary to urgently investigate PTs' perception and preparation for using these advanced technologies. Understanding PTs' perception would help in maximizing confidence in, and enhancing comfort regarding, the use of 21st century advanced technologies in physical therapy practices.

Objectives

Because of the fast pace of innovation in AI and digital technologies, it is impossible to ignore the current debate about the importance of these technologies in clinical practice, especially in rehabilitation. However, we have to first set the stage for this digital revolution by ascertaining PTs' knowledge and attitudes regarding this new era of health and rehabilitation practices. Although previous research has identified the various applications of AI in health care and rehabilitation [16], little has been investigated about PTs' knowledge and attitudes regarding AI applications. Therefore, the aim of this study was to explore PTs' understanding of the AI technologies used in health care and rehabilitation. In addition, this study assessed the relationship between PTs' knowledge and multiple demographic variables, including sex, educational qualification, years of experience, workplace setting, and number of AI applications at work. The results of this study would help in filling the gap in current research recommendations as well as academic and clinical practices.

Methods

Participants

PTs were invited to voluntarily participate in the study. Only participants working in Saudi Arabia as licensed PTs could participate. In March 2021, a survey link was created using Google Forms (Google LLC). In the prefatory section of the survey, a brief description was provided to inform the participants about the goal of the study and to confirm the confidentiality and anonymity of their data. To obtain informed consent, a question about the participants' agreement to participate in the study was placed at the beginning of the survey.

Ethics Approval

Institutional review board approval was received from the Center of Excellence in Genomic Medicine Research (14-CEGMR-Bioeth-2021), approved by the National Committee of Bioethics, Jeddah, Saudi Arabia (KACST: HA-02-J-003).

Instrument

The survey was developed through deep searching in the literature [2,12] and feedback from physical therapy experts. The face validity and content validity of the survey [17,18] were established by inviting 8 PTs who were experts in the field of rehabilitation and survey studies to review and rate each item of the survey for its appropriateness, clarity, ordering, and construct. Next, each expert's comments were reviewed by the principal investigator to improve the quality of the survey questions and establish the content validity of the survey based upon 80% agreement of the experts' feedback. The content validity index was 0.8 for the whole survey; however, the content validity index was between 0.8 and 1 for each item.

The survey, which was divided into 4 sections, consisted of 20 questions. The first section of the survey asked about the demographic characteristics of the participants to determine the sample age, sex, years of experience, educational qualification,

number of AI applications at work, and subspecialization. The second section asked about participants' knowledge about AI in the field of health care and rehabilitation. The third section sought participants' opinions regarding the advantages and uses of AI as well as its impact on the future of rehabilitation. The final section concerned the ethical implications of using AI and participants' willingness to explore the AI field. The answers to the survey questions were assessed using yes or no questions and a 5-point Likert scale (ranging from strongly agree to strongly disagree).

In this study, we investigated whether PTs' knowledge and attitudes regarding AI medical applications differed depending on the respondent's sex, years of experience, educational qualification, employment sector, and number of AI applications at work. For this study, years of experience were categorized as >10 years or <10 years. The work sector categories were nonacademic or academic. Educational qualification categories were undergraduate or postgraduate (master's degree and PhD), whereas the number of AI applications at work was categorized as no AI application or at least one AI application.

Procedures

The electronic open survey was distributed using social media, including WhatsApp, Facebook groups, and Twitter. In addition, contact was made with PTs via email with a request to forward the survey to other PTs if they knew one. The survey was open from March 2021 to May 2021. Before distributing the survey, the minimum sample size was calculated using G*Power (version 3.1; Heinrich Heine University) to achieve a power of 0.80. In the G*Power software, a logistic regression test was conducted for a priori power calculation with an odds ratio (OR) of 1.5 and significance level of .05. The minimum sample needed to achieve a power of 0.80 was 280 for our study. This indicates that the sample size attained in this study (n=317) was sufficient to detect an effect. The report of this study has been written according to the Checklist for Reporting Results of Internet E-Surveys guidelines [19].

Statistical Analysis

After the data were collected, they were coded and entered into a spreadsheet using Microsoft Excel 2016. Data were analyzed using SPSS software (version 28.0; IBM Corp). Descriptive statistics were used to describe the sample's demographic characteristics in frequencies and percentages. Chi-square tests and binary logistic regression analysis were used to investigate the differences in PTs' perceptions regarding AI applications in health care and rehabilitation based on demographic characteristics. A P value of ≤.05 was considered statistically significant.

Results

Descriptive Statistics

A total of 317 PTs from different workplace settings participated in the study. The mean age of the participants was 33.38 (SD 7.31) years. With regard to the respondents' sex, 52.4% (166/317) of the participants were male, and 47.6% (151/317) were female. Most (243/317, 76.7%) of the participants were working in nonacademic sectors, mainly at outpatient clinics and hospitals. Nearly half (152/317, 47.9%) of the participants were general PTs. The majority (193/317, 60.9%) of the respondents reported that they had not come across any AI application at work. Only a few (11/317, 3.5%) of the participants had been exposed to AI applications at work >4 times. Of the 317 respondents, 137 (43.2%) reported that they obtained information about AI primarily from social media, whereas 114 (36%) stated that they obtained information about AI primarily from articles and journals. Detailed demographic characteristics of the participants are presented in Table 1.

Table 1. Demographic characteristics of the participants, exposure to artificial intelligence (AI) applications at work, and sources of AI information (N=317).

Characteristics	Values
Age (years), mean (SD); median (range)	33.38 (7.31); 33 (22-63)
Sex, n (%)	
Male	166 (52.4)
Female	151 (47.6)
Employment sector, n (%)	
Academic	74 (23.3)
Nonacademic	243 (76.7)
Educational qualification, n (%)	
Undergraduate degree	188 (59.3)
Postgraduate degree	129 (40.7)
Subspecialty, n (%)	
Cardiorespiratory	15 (4.7)
General	152 (47.9)
Geriatrics	8 (2.5)
Musculoskeletal and sports	102 (32.2)
Neurorehabilitation	23 (7.3)
Pediatrics rehabilitation	17 (5.4)
AI applications at work, n (%)	
0	194 (61.2)
1	56 (17.7)
2 to 4	56 (17.7)
>4	11 (3.5)
Source of AI information (multiple responses), n (%)	
Social media	137 (43.2)
Traditional media	50 (15.8)
Colleagues or friends	97 (30.6)
Class lectures	80 (25.2)
Articles or journals	114 (36)
Workshops	44 (13.9)
Work	37 (11.7)
Web-based courses	29 (9.1)
No prior information	36 (11.4)

Knowledge About AI

Table 2 shows the results of the binary logistic regression analysis to find the statistically significant differences in the PTs' AI knowledge—general, health care, and rehabilitation—based on the demographic variables.

Table 2. Results of logistic regression analysis to assess the factors associated with artificial intelligence (AI; N=317).

Variable	B	95% CI for B	SE for B	β	*P* value
Knowledge about AI in general					
Sex					
Constant	1.48	N/A[a]	0.21	N/A	N/A
Male	1.38	1.81-8.73	0.40	3.97	.001
Female	Reference	N/A	N/A	N/A	N/A
Employment sector					
Constant	1.55	N/A	0.31	N/A	N/A
Nonacademic	0.67	0.94-4.04	0.37	4.70	.07
Academic	Reference	N/A	N/A	N/A	N/A
Experience (years)					
Constant	1.79	N/A	0.21	N/A	N/A
>10	0.69	0.93-4.28	0.39	2.00	.08
<10	Reference	N/A	N/A	N/A	N/A
Educational qualification					
Constant	1.838	N/A	0.21	N/A	N/A
Postgraduate degree	0.54	0.82-3.62	0.38	1.72	.15
Undergraduate degree	Reference	N/A	N/A	N/A	N/A
AI at workplace					
Constant	1.62	N/A	0.19	N/A	N/A
≥1	1.54	1.76-12.32	N/A	4.66	.002
0	Reference	N/A	N/A	N/A	N/A
Knowledge about AI in health care					
Sex					
Constant	0.50	N/A	0.17	N/A	N/A
Male	1.38	2.28-6.92	0.28	3.97	<.001
Female	Reference	N/A	N/A	N/A	N/A
Employment sector					
Constant	0.81	N/A	0.25	N/A	N/A
Nonacademic	0.41	0.85-2.68	0.30	1.51	.16
Academic	Reference	N/A	N/A	N/A	N/A
Experience (years)					
Constant	0.83	N/A	0.16	N/A	N/A
>10	0.75	1.22-3.68	0.28	2.12	.008
<10	Reference	N/A	N/A	N/A	N/A
Educational qualification					
Constant	0.78	N/A	0.16	N/A	N/A
Postgraduate degree	0.91	1.41-4.40	0.30	2.50	.002
Undergraduate degree	Reference	N/A	N/A	N/A	N/A
AI at workplace					
Constant	0.57	N/A	0.15	N/A	N/A
≥1	1.96	3.41-14.97	0.37	7.15	<.001
0	Reference	N/A	N/A	N/A	N/A

Variable	B	95% CI for B	SE for B	β	*P* value
Knowledge about AI in rehabilitation					
Sex					
Constant	0.04	N/A	0.16	N/A	N/A
Male	0.89	1.53-3.87	0.24	2.43	<.001
Female	Reference	N/A	N/A	N/A	N/A
Employment sector					
Constant	0.27	N/A	0.24	N/A	N/A
Nonacademic	0.28	0.78-2.24	0.27	1.32	.31
Academic	Reference	N/A	N/A	N/A	N/A
Experience (years)					
Constant	0.26	N/A	0.15	N/A	N/A
>10	0.58	1.11-2.87	0.24	1.79	.02
<10	Reference	N/A	N/A	N/A	N/A
Educational qualification					
Constant	0.28	N/A	0.15	N/A	N/A
Postgraduate degree	0.52	1.05-2.70	0.24	1.68	.03
Undergraduate degree	Reference	N/A	N/A	N/A	N/A
AI at workplace					
Constant	–0.21	N/A	0.14	N/A	N/A
≥1	2.26	5.13-17.87	0.32	9.57	<.001
0	Reference	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

Knowledge About AI in General

Overview

Of the 317 PTs, 280 (88.3%) indicated that they had knowledge about AI in general. The data indicated that there were no statistically significant differences in AI general knowledge by employment sector, experience, or qualification. However, there was a significant difference by sex (OR 3.97, 95% CI 1.81-8.73; *P*=.001); that is, the male PTs were 3.97 times more knowledgeable about AI in general than the female PTs. In addition, this study found that the number of AI applications at work was a statistically significant predictor of AI general knowledge among PTs (OR 4.66, 95% CI 1.76-12.32; *P*=.002).

Knowledge About AI in Health Care

Of the 317 PTs, 238 (75.1%) indicated that they had knowledge about AI in the field of health care. In this study, employment sector was not a significant predictor of knowledge about AI in health care among the PTs. However, there was a significant difference in AI knowledge based on sex (OR 3.96, 95% CI 2.28-6.92; *P*<.001). Compared with the female PTs, the male PTs were 3.96 times more likely to be familiar with AI applications. Participants who had >10 years of experience were 2.12 times more knowledgeable about AI applications than those with less experience (*P*=.008). In addition, there was a significant difference in knowledge about AI in health care based on educational qualification (OR 2.5, 95% CI 1.41-4.40; *P*=.002). The results indicated that PTs with an undergraduate degree were 2.5 times less knowledgeable about AI applications than those with a postgraduate degree. Furthermore, the findings revealed that PTs who had experience of working with at least one AI application were 7.15 times more knowledgeable about AI health care technologies than those who had no experience of working with AI applications (*P*<.001).

Knowledge About AI in Rehabilitation

Of the 317 PTs, 121 (38.2%) reported that they had knowledge about AI applications in the rehabilitation field. The results showed that there was a statistically significant difference in PTs' knowledge regarding AI in rehabilitation based on sex (OR 2.43, 95% CI 1.53-3.87; *P*<.001): the male PTs were 2.43 times more knowledgeable about AI use in rehabilitation than the female PTs. In addition, experience and educational qualification were significant predictors of knowledge about AI applications in rehabilitation among the PTs: OR 1.79, 95% CI 1.11-2.87; *P*=.02, and OR 1.68, 95% CI 1.05-2.70; *P*=.03, respectively. Moreover, the number of AI applications at work was a significant predictor of AI knowledge in rehabilitation (*P*<.001). The results implied that having worked with at least one AI application increases AI knowledge by 9.57 times compared with having no practical experience at work.

Attitudes Regarding Advantages of AI

Using a 5-point Likert scale, participants indicated their level of agreement regarding three advantages of using AI applications in health care and rehabilitation: reducing therapist workload, prevention of diseases, and facilitating patient care. The participants' levels of agreement (frequencies and percentages) regarding the advantages of using AI in clinical practice are detailed in Table 3.

Table 3. Participants' attitudes regarding the advantages of using artificial intelligence (AI) in clinical practice (N=317).

Advantages of using AI in clinical practice and variables	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Reducing therapist workload, n (%)					
Sex					
Male	67 (21.1)	59 (18.6)	34 (10.7)	5 (1.6)	1 (0.3)
Female	41 (12.9)	68 (21.5)	36 (11.4)	6 (1.9)	0 (0)
Employment sector					
Academic	23 (7.3)	35 (11)	12 (3.8)	3 (0.9)	1 (0.3)
Nonacademic	85 (26.8)	92 (29)	58 (18.3)	8 (2.5)	0 (0)
Experience (years)					
>10	52 (16.4)	37 (11.7)	32 (10.1)	7 (2.2)	1 (0.3)
<10	56 (17.7)	90 (28.4)	38 (12)	4 (1.3)	0 (0)
Educational qualification					
Postgraduate degree	55 (17.4)	45 (14.2)	23 (7.3)	5 (1.6)	1 (0.3)
Undergraduate degree	53 (16.7)	82 (25.9)	47 (14.8)	6 (1.9)	0 (0)
Facilitating patient care, n (%)					
Sex					
Male	65 (20.5)	77 (24.3)	20 (6.3)	2 (0.6)	2 (0.6)
Female	43 (13.6)	78 (24.6)	25 (7.9)	5 (1.6)	0 (0)
Employment sector					
Academic	26 (8.2)	36 (11.4)	9 (2.8)	3 (0.9)	0 (0)
Nonacademic	82 (25.9)	119 (37.5)	36 (11.4)	4 (1.3)	2 (0.6)
Experience (years)					
>10	44 (13.9)	61 (19.2)	21 (6.6)	3 (0.9)	0 (0)
<10	64 (20.2)	94 (29.7)	24 (7.6)	4 (1.3)	2 (0.6)
Educational qualification					
Postgraduate degree	42 (13.2)	68 (21.5)	14 (4.4)	3 (0.9)	2 (0.6)
Undergraduate degree	66 (20.8)	87 (27.4)	31 (9.8)	4 (1.3)	0 (0)
Prevention of diseases, n (%)					
Sex					
Male	51 (16.1)	49 (15.5)	53 (16.7)	10 (3.2)	3 (0.9)
Female	31 (9.8)	52 (16.4)	53 (16.7)	13 (4.1)	2 (0.6)
Employment sector					
Academic	27 (8.5)	25 (7.9)	17 (5.4)	5 (1.6)	0 (0)
Nonacademic	55 (17.4)	76 (24)	89 (28.1)	18 (5.7)	5 (1.6)
Experience (years)					
>10	32 (10.1)	35 (11)	51 (16.1)	8 (2.5)	3 (0.9)
<10	50 (15.8)	66 (20.8)	55 (17.4)	15 (4.7)	2 (0.6)
Educational qualification					
Postgraduate degree	39 (12.3)	35 (11)	41 (12.9)	11 (3.5)	3 (0.9)
Undergraduate degree	43 (13.6)	66 (20.8)	65 (20.5)	12 (3.8)	2 (0.6)

Reducing Therapist Workload

More male PTs (126/317, 39.7%) than female PTs (121/317, 38.2%) agreed or strongly agreed that using AI reduces therapist workload. Moreover, a high percentage of the nonacademic participants (177/243, 72.8%) agreed or strongly agreed that using AI reduces the workload of PTs in clinical practice, whereas 77.7% (146/188) of the participants with <10 years of experience agreed or strongly agreed that using AI reduces the workload in PTs' clinical practice. However, on the basis of educational qualification, a few participants (12/317, 3.8%) disagreed or strongly disagreed that AI is useful in reducing PTs' workload.

Facilitating Patient Care

Of the 317 participants, more male PTs (n=142, 44.8%) reported their agreement that AI applications have advantages in facilitating patient care than female PTs (n=121, 38.2%). Of the 243 participants working in the nonacademic sector, most (n=201, 82.7%) agreed or strongly agreed that AI technologies can facilitate patient care in clinical practice. With regard to educational qualification, 81.4% (153/188) of the participants who had an undergraduate degree agreed or strongly agreed that using AI facilitated patient care in clinical settings. On the basis of years of experience, only 2.8% (9/317) of the participants disagreed or strongly disagreed that AI would be useful in facilitating patient care.

Prevention of Diseases

Of the 317 participants, more male PTs (n=100, 31.5%) reported positive attitudes regarding the advantage of AI technologies in preventing diseases than female PTs (n=83, 26.2%). In addition, of the 243 participants working in nonacademic sectors, 131 (53.9%) indicated that AI applications have a role in preventing diseases, whereas of the 74 participants working in academic organizations, 52 (70%) indicated that AI applications have a role in preventing diseases. In addition, the study results showed that of the 188 PTs with <10 years of experience, 116 (61.7%) had positive attitudes regarding using AI technologies to prevent diseases. Furthermore, participants with an undergraduate degree had a slightly higher level of agreement regarding the usefulness of AI technologies in preventing diseases than those with a postgraduate degree (109/188, 58%, vs 74/129, 57.3%, respectively).

Uses of AI

Participants were asked to indicate their level of agreement regarding five aspects of the uses of AI: disease prediction, goal setting, assistive technologies, diagnostic tool, and education enhancement. Table 4 shows in detail the attitudes of the PTs regarding the uses of AI in clinical settings.

Table 4. Participants' attitudes regarding the uses of artificial intelligence (AI) in clinical practice (N=317).

Uses of AI in clinical practice and variables	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Disease prediction, n (%)					
Sex					
Male	41 (12.9)	68 (21.5)	49 (15.5)	5 (1.6)	3 (0.9)
Female	20 (6.3)	53 (16.7)	63 (19.9)	12 (3.8)	3 (0.9)
Employment sector					
Academic	17 (5.4)	28 (8.8)	27 (8.5)	1 (0.3)	1 (0.3)
Nonacademic	44 (13.9)	93 (29.3)	85 (26.8)	16 (5)	5 (1.6)
Experience (years)					
>10	21 (6.6)	63 (19.9)	39 (12.3)	5 (1.6)	1 (0.3)
<10	40 (12.6)	58 (18.3)	73 (23)	12 (3.8)	5 (1.6)
Educational qualification					
Postgraduate degree	32 (10.1)	48 (15.1)	41 (12.9)	5 (1.6)	3 (0.9)
Undergraduate degree	29 (9.1)	73 (23)	71 (22.4)	12 (3.8)	3 (0.9)
Goal setting, n (%)					
Sex					
Male	44 (13.9)	74 (23.3)	22 (6.9)	24 (7.6)	2 (0.6)
Female	23 (7.3)	63 (19.9)	48 (15.1)	14 (4.4)	3 (0.9)
Employment sector					
Academic	20 (6.3)	32 (10.1)	20 (6.3)	2 (0.6)	0 (0)
Nonacademic	47 (14.8)	105 (33.1)	50 (15.8)	36 (11.4)	5 (1.6)
Experience (years)					
>10	26 (8.2)	60 (18.9)	24 (7.6)	19 (6)	0 (0)
<10	41 (12.9)	77 (24.3)	46 (14.5)	9 (6)	5 (1.6)
Educational qualification					
Postgraduate degree	32 (10.1)	58 (18.3)	15 (4.7)	22 (6.9)	2 (0.6)
Undergraduate degree	35 (11)	79 (24.9)	55 (17.4)	16 (5)	3 (0.9)
Assistive technologies, n (%)					
Sex					
Male	64 (20.2)	88 (27.8)	12 (3.8)	2 (0.6)	0 (0)
Female	55 (17.4)	70 (22.1)	25 (7.9)	1 (0.3)	0 (0)
Employment sector					
Academic	22 (6.9)	41 (12.9)	10 (3.2)	1 (0.3)	0 (0)
Nonacademic	97 (30.6)	117 (36.9)	27 (8.5)	2 (0.6)	0 (0)
Experience (years)					
>10	40 (12.6)	75 (23.7)	13 (4.1)	1 (0.3)	0 (0)
<10	79 (24.9)	83 (26.2)	24 (7.6)	2 (0.6)	0 (0)
Educational qualification					
Postgraduate degree	42 (13.2)	78 (24.6)	8 (2.5)	1 (0.3)	0 (0)
Undergraduate degree	77 (24.3)	80 (25.2)	29 (9.1)	2 (0.6)	0 (0)
Diagnostic tool, n (%)					
Sex					
Male	49 (15.5)	78 (24.6)	23 (7.3)	11 (3.5)	5 (1.6)

Uses of AI in clinical practice and variables	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Female	32 (10.1)	64 (20.2)	43 (13.6)	10 (3.2)	2 (0.6)
Employment sector					
Academic	24 (7.6)	33 (10.4)	11 (3.5)	6 (1.9)	0 (0)
Nonacademic	57 (18)	109 (34.4)	55 (17.4)	15 (4.7)	7 (2.2)
Experience (years)					
>10	31 (9.8)	73 (23)	22 (6.6)	3 (0.9)	0 (0)
<10	50 (15.8)	69 (21.8)	44 (13.9)	18 (5.7)	7 (2.2)
Educational qualification					
Postgraduate degree	36 (11.4)	67 (21.1)	14 (4.4)	9 (2.8)	3 (0.9)
Undergraduate degree	45 (14.2)	75 (23.7)	52 (16.4)	12 (3.8)	4 (1.3)
Education enhancement, n (%)					
Sex					
Male	75 (23.7)	73 (23.1)	12 (3.8)	4 (1.3)	2 (0.6)
Female	54 (17)	66 (20.8)	29 (9.1)	2 (0.6)	0 (0)
Employment sector					
Academic	33 (10.4)	34 (10.7)	6 (1.9)	1 (0.3)	0 (0)
Nonacademic	96 (30.3)	105 (33.1)	35 (11)	5 (1.6)	2 (0.6)
Experience (years)					
>10	47 (14.8)	71 (22.4)	9 (2.8)	2 (0.6)	0 (0)
<10	82 (25.9)	68 (21.5)	32 (10.1)	4 (1.3)	2 (0.6)
Educational qualification					
Postgraduate degree	59 (18.6)	62 (19.6)	5 (1.6)	1 (0.3)	2 (0.6)
Undergraduate degree	70 (22.1)	77 (24.3)	36 (11.4)	5 (1.6)	0 (0)

Disease Prediction

Of the 166 male PTs, 109 (65.7%) agreed or strongly agreed that disease prediction is one of the uses of AI applications in clinical settings. In addition, 56.4% (137/243) of the participants working in nonacademic settings reported their agreement regarding using AI technologies in disease prediction. Of the 188 participants with <10 years of experience, 98 (52.1%) agreed or strongly agreed that disease prediction can be provided by AI technologies. However, on the basis of educational qualification, 7.2% (23/317) of the participants disagreed or strongly disagreed that AI could be used for predicting diseases.

Goal Setting

Of the 166 male participants, 118 (71.1%) agreed or strongly agreed regarding using AI applications for goal setting, whereas only 26 (15.7%) disagreed or strongly disagreed with the statement. Of the 74 participants working in academic organizations, only 2 (3%) disagreed or strongly disagreed that AI can be used for goal-setting purposes. With regard to years of experience, the majority (204/317, 64.4%) of the participants agreed or strongly agreed that goal setting could be facilitated by AI technologies. Similarly, on the basis of educational qualification, the majority (204/317, 64.4%) of the PTs agreed or strongly agreed that goal setting could be facilitated by AI.

Assistive Technologies

Of the 317 participants, 277 (87.4%) agreed or strongly agreed that AI applications can be used as assistive technologies in health care and rehabilitation. However, the male PTs (152/166, 91.6%) had a higher level of agreement than the female PTs (125/151, 82.8%). The results indicated that of the 243 participants working in the nonacademic sector, 214 (88.1%) agreed or strongly agreed that AI applications are among the assistive technologies used in the medical field. On the basis of experience and educational qualification, very few (3/317, 0.9%, in each category) of the participants disagreed about using AI applications as assistive technologies in health care.

Diagnostic Tool

Of the 166 male PTs, 127 (76.5%) agreed or strongly agreed that AI applications can be used to determine patients' diagnoses. The majority (166/243, 68.3%) of the participants working in the nonacademic sector indicated that AI may help clinicians in providing medical diagnoses. In addition, we found that 63.8% (120/188) of the PTs with an undergraduate degree agreed or strongly agreed that AI technologies could be used for diagnostic purposes compared with 57.9% (73/129) of those with a postgraduate degree.

Education Enhancement

Of the 166 male participants, 148 (89.2%) agreed or strongly agreed about using AI technologies to enhance education among health care providers. In addition, our results revealed that of the 243 participants working in the nonacademic sector, 210 (86.4%) highly supported using AI technologies for education enhancement in the medical field. On the basis of experience and education, very few (2/317, 0.6%, in each category) of the participants strongly disagreed that AI has a role in enhancing the educational background of practitioners.

Impact of AI

Participants were asked to indicate their level of agreement regarding 3 impacts of using AI technologies in health care and rehabilitation. The detailed results are presented in Table 5.

Table 5. Participants' attitudes regarding the impact of artificial intelligence (AI) on the rehabilitation field (N=317).

The impact of AI and variables	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Reducing human resources, n (%)					
Sex					
Male	57 (18)	49 (15.5)	23 (10.1)	25 (7.9)	3 (0.9)
Female	27 (8.5)	55 (17.4)	50 (15.8)	15 (4.7)	4 (1.3)
Employment sector					
Academic	15 (4.7)	30 (9.5)	20 (6.3)	8 (2.5)	1 (0.3)
Nonacademic	69 (21.8)	74 (23.3)	62 (19.6)	32 (10.1)	6 (1.9)
Experience (years)					
>10	43 (13.6)	41 (12.9)	25 (7.9)	19 (6)	1 (0.3)
<10	41 (12.9)	63 (19.9)	57 (18)	21 (6.6)	6 (1.9)
Educational qualification					
Postgraduate degree	50 (15.8)	34 (10.7)	26 (8.2)	17 (5.4)	2 (0.6)
Undergraduate degree	34 (10.7)	70 (22.1)	56 (17.7)	23 (7.3)	5 (1.6)
Increasing productivity, n (%)					
Sex					
Male	52 (16.4)	86 (27.1)	23 (7.3)	3 (0.9)	2 (0.6)
Female	40 (12.6)	70 (22.1)	34 (10.7)	5 (1.6)	2 (0.6)
Employment sector					
Academic	20 (6.3)	37 (11.7)	14 (4.4)	3 (0.9)	0 (0)
Nonacademic	72 (22.7)	119 (37.5)	43 (13.6)	5 (1.6)	4 (1.3)
Experience (years)					
>10	32 (10.1)	73 (23)	21 (6.6)	3 (0.9)	0 (0)
<10	60 (18.9)	83 (26.2)	36 (11.4)	5 (1.6)	1 (1.3)
Educational qualification					
Postgraduate degree	37 (11.7)	66 (20.8)	22 (6.9)	2 (0.6)	2 (0.6)
Undergraduate degree	55 (17.4)	90 (28.4)	35 (11)	6 (1.9)	2 (0.6)
Improving patients' quality of life, n (%)					
Sex					
Male	60 (18.9)	65 (20.5)	34 (10.7)	5 (1.6)	2 (0.6)
Female	42 (13.2)	65 (20.5)	36 (11.4)	5 (1.6)	3 (0.9)
Employment sector					
Academic	24 (7.6)	33 (10.4)	14 (4.4)	3 (0.9)	0 (0)
Nonacademic	78 (24.6)	97 (30.6)	56 (17.7)	7 (2.2)	5 (1.6)
Experience (years)					
>10	38 (12)	52 (16.4)	33 (10.4)	6 (1.9)	0 (0)
<10	64 (20.2)	78 (24.6)	37 (11.7)	4 (1.3)	5 (1.6)
Educational qualification					
Postgraduate degree	44 (13.9)	48 (15.1)	30 (9.5)	5 (1.6)	2 (0.6)
Undergraduate degree	58 (18.3)	82 (25.9)	40 (12.6)	5 (1.6)	3 (0.9)

Reducing Human Resources

Of the 166 male participants, 106 (63.9%) agreed that AI use has an impact on human resource reduction. In addition, we found that 58.8% (143/243) of the participants working in the nonacademic sector were highly of the opinion that the use of AI technologies may result in the reduction of human resources in the clinical field. Of the 188 PTs with <10 years of experience, 104 (55.3%) agreed or strongly agreed that AI use may result in human resource reduction. The results also showed that the participants with a postgraduate degree had a higher level of strong agreement about the impact of AI use on human resource reduction than those with an undergraduate degree (50/129, 38.8%, vs 34/188, 18.1%, respectively).

Increasing Productivity

The results showed that the majority (248/317, 78.2%) of the participants agreed or strongly agreed that work productivity could be increased by implementing AI in health care. Of the 243 participants working in the nonacademic sector, 191 (78.6%) agreed or strongly agreed that AI use could increase productivity. Very few (3/129, 2.3%) of the participants with >10 years of experience disagreed that an increase in productivity could be achieved by using AI technologies in health care. The PTs with an undergraduate degree had a slightly lower level of agreement on the role of AI in improving work productivity than those with a postgraduate degree (145/188, 77.1%, vs 103/129, 79.8%, respectively).

Improving Patients' Quality of Life

The study found that more male participants (125/166, 75.3%) than female participants (107/151, 70.9%) agreed or strongly agreed that patients' quality of life can be improved by using AI technologies in health care and rehabilitation. In addition, the study results indicated that 72% (175/243) of the participants working in the nonacademic sector significantly agreed that AI has a positive impact on patients' quality of life. Furthermore, 75.5% (142/188) of the participants with <10 years of experience agreed or strongly agreed that using AI has a positive impact on patients' quality of life.

Ethical Implications of Using AI and Willingness to Explore the AI Field

The study investigated PTs' ethical concerns that might arise when implementing AI in health care and rehabilitation settings. Nearly half (144/317, 45.4%) of the participants expressed concerns about the inability of AI applications to sympathize with human beings or understand the complexity of the human experience, whereas 42.9% (136/317) were concerned about the inability of AI applications used in health care to provide a judgment in unpredicted situations that are beyond the scope of the AI program. In addition, a few (36/317, 11.4%) of the respondents stated that they were concerned about AI developers not being from the medical field or having minimal experience in medical or clinical practice.

In addition, in response to the question "If the clinician's judgment clashed with that of the AI application, which one should be trusted?" only 6% (19/317) of the participants stated that the AI application's decision should be trusted. Most (262/317, 82.6%) of the PTs reported that the clinician's judgment should be preferred over that of the AI application, whereas 11% (35/317) of the respondents expressed a preference for abiding by the patient's choice when the clinician's reasoning conflicted with the AI application's decision. However, in response to a question about whether AI courses should be included in rehabilitation curricula, 71.9% (228/317) of the PTs responded in the positive.

Discussion

Principal Findings

The main purpose of this study was to obtain a snapshot of the overall perceptions and attitudes of PTs regarding AI applications in health care and rehabilitation. This study assessed the relationships among multiple factors, including sex, experience, employment sector, and educational qualification. To the best of our knowledge, this is the first study that examines PTs' thoughts and opinions regarding AI technologies and their relationships with multiple explanatory variables. The study findings might add to the existing knowledge regarding why it is important to enhance PTs' awareness of the advantages and uses of AI technologies in clinical practice.

In this study, it was found that the majority (health care: 238/317, 75.1%, and rehabilitation: 121/317, 38.2%) of the participants had moderate knowledge about AI in health care and rehabilitation. Most (196/317, 61.8%) of the respondents stated that they had not heard about AI applications in rehabilitation. The results were consistent with those of a study that was conducted in Canada to explore the perceptions of oncologists, physicists, and radiation therapists about AI, which reported moderate knowledge about AI applications in medicine [20]. In addition, similar findings were reported in an Australian study that highlighted the average knowledge about the impact of AI among different health care professions [21]. Surprisingly, the majority (194/317, 61.2%) of the respondents in this study reported that they had not come across any AI applications at their workplace. Although AI technologies have been a focus of medical research, real-world clinical practice still faces obstacles when it comes to implementing AI. To successfully implement AI technologies in rehabilitation, PTs need to have prior knowledge, practical experience, confidence, and acceptance of AI technologies. This study did not investigate the barriers to AI implementation in clinical practice; therefore, research could be conducted in the future to support this study's findings.

Generally, the male participants reported having more knowledge and more positive attitudes regarding AI applications than the female participants. Similar findings were reported by Santos et al [22] who found that male students were more interested than female students in AI and robotics. Moreover, most (223/317, 70.3%) of the PTs expressed the view that AI applications would have an impact on health care and rehabilitation practices. However, participants with <10 years of experience were more likely to believe that AI would have an impact on clinical practice. This was consistent with the results of a previous study by Scheetz et al [21], which indicated that health care practitioners with fewer years of clinical practice, including ophthalmologists, radiation oncologists, and

dermatologists, agreed that AI would have an impact on the workforce. The reasons behind this have not been investigated previously. However, it is possible, as noted in the results, that clinicians with more experience have less confidence in AI.

In this study, most (218/317, 68.8%) of the participants stated that they believed that AI would reduce PTs' workload and increase their productivity. This finding was similar to those of studies of AI use among other clinicians [20,23]. However, employment sector was one of the explanatory factors in this study. We found that there was a statistically significant difference in the PTs' responses based on their primary workplace. Participants working in the nonacademic sector were more likely to accept AI applications than PTs who worked in the academic sector. There are no prior studies on the differences in PTs' knowledge and attitudes regarding AI based on their primary practice setting; therefore, the explanation is not clear, and more research needs to be conducted in this area. It is essential to have a better understanding of physical therapy educators' knowledge and practical experience of, as well as confidence in, AI technologies because they are among the facilitators who would increase the acceptance of AI applications by future PTs.

Incorporating AI technologies in the physical therapy core curriculum would help to smoothen future PTs' engagement with the new era of intelligent technologies in rehabilitation practices. Future PTs need to be mentally prepared to explore, understand, and apply the algorithms of AI applications in their practice. In this study, 71.9% (228/317) of the participants indicated that AI courses should be incorporated in the academic curriculum. Previous studies also suggested integrating different courses related to AI into undergraduate and postgraduate programs such as data science, deep learning, and behavioral science, which may help clinicians to understand and apply AI in their medical practice [6,24].

In addition, the results of this study indicated that only 6% (19/317) of the PTs think that the AI application's decision should be preferred over that of the clinician, whereas the majority (262/317, 82.6%) of the PTs stated that they would abide by the clinician's decision. A similar result was reported in the study by Oh et al [25], who found that the majority of the doctors would favor trusting their own opinion over that of the AI application when there was a difference of opinion. In this study, the results indicated that there is insufficient information about PTs' knowledge of, and experience with, AI applications, especially in the rehabilitation field. This paper promotes the necessity for more research to be conducted to increase the knowledge and practical experience of PTs regarding AI applications.

This study includes some limitations. First, because the survey was self-administered, there is a possibility of some bias regarding the PTs' responses. In addition, the results cannot be generalized to other health care professionals because this study was limited to PTs. In this study, an electronic survey was used to collect the data, and this may have led to sample selection bias. Other sampling strategies could be used in the future to reach out to a more representative sample of PTs. In physical therapy research, AI applications are being developed rapidly, but a very limited number of AI techniques are being implemented and translated into physical therapy practices. This study's results indicate low-to-average AI knowledge among PTs and positive attitudes regarding the different advantages, uses, and impacts of AI use. However, action is required to translate AI technologies from research into actual clinical practice.

Conclusions

The use of AI technologies is growing rapidly in health care and rehabilitation. Thus, there is a need to increase PTs' awareness of various AI applications in rehabilitation to provide competent patient care facilities. The results of this study indicate that being a man, having >10 years of experience, and having a postgraduate degree are the anticipated PT criteria that increase AI knowledge and adoption levels. In addition, the results highlighted the importance of promoting evidence-based knowledge translation, particularly with regard to AI technologies, among PTs. However, to successfully implement AI in the rehabilitation field, further research on both physical therapy clinicians and patient expectations should be conducted.

Acknowledgments

This work was funded by the Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2022R267), Princess Nourah bint Abdulrahman University (PO Box 84428, Riyadh 11671, Saudi Arabia).

Conflicts of Interest

None declared.

References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
2. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. Front Artif Intell 2020;3:578983 [FREE Full text] [doi: 10.3389/frai.2020.578983] [Medline: 33733219]
3. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018 Nov;15(11):e1002686 [FREE Full text] [doi: 10.1371/journal.pmed.1002686] [Medline: 30457988]

4. Hwang J, Jung Y, Cho B, Heo M. An overview of deep learning in the field of dentistry. Imaging Sci Dent 2019 Mar;49(1):1-7 [FREE Full text] [doi: 10.5624/isd.2019.49.1.1] [Medline: 30941282]
5. 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 Feb 02;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
6. Rowe M, Nicholls DA, Shaw J. How to replace a physiotherapist: artificial intelligence and the redistribution of expertise. Physiother Theory Pract 2021 Jun 03:1-9. [doi: 10.1080/09593985.2021.1934924] [Medline: 34081573]
7. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7:e7702 [FREE Full text] [doi: 10.7717/peerj.7702] [Medline: 31592346]
8. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014 Jan;16(1):441. [doi: 10.1007/s11886-013-0441-8] [Medline: 24338557]
9. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol 2013 Sep;201(3):611-617. [doi: 10.2214/AJR.12.10375] [Medline: 23971454]
10. Vélez-Guerrero MA, Callejas-Cuervo M, Mazzoleni S. Artificial intelligence-based wearable robotic exoskeletons for upper limb rehabilitation: a review. Sensors (Basel) 2021 Mar 18;21(6):2146 [FREE Full text] [doi: 10.3390/s21062146] [Medline: 33803911]
11. Ravali RS, Vijayakumar TM, Santhana Lakshmi K, Mavaluru D, Reddy LV, Retnadhas M, et al. A systematic review of artificial intelligence for pediatric physiotherapy practice: past, present, and future. Neurosci Informatics 2022 Dec;2(4):100045. [doi: 10.1016/j.neuri.2022.100045]
12. Brougham D, Haar J. Smart technology, artificial intelligence, robotics, and algorithms (STARA): employees' perceptions of our future workplace. J Manage Organization 2017 Jan 24;24(2):239-257. [doi: 10.1017/jmo.2016.55]
13. Anderson D. Am J Phys Med Rehabil 2019 Mar 5;98(11):e128-e129. [doi: 10.1097/phm.0000000000001171]
14. Ramanandi V. Role and scope of artificial intelligence in physiotherapy: a scientific review of literature. Int J Adv Sci Res 2021;6(1):11-14.
15. Ye C, Li J, Hao S, Liu M, Jin H, Zheng L, et al. Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. Int J Med Inform 2020 May;137:104105. [doi: 10.1016/j.ijmedinf.2020.104105] [Medline: 32193089]
16. Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. Musculoskelet Sci Pract 2019 Feb;39:164-169. [doi: 10.1016/j.msksp.2018.11.012] [Medline: 30502096]
17. Yusoff MS. ABC of response process validation and face validity index calculation. Educ Med J 2019 Oct 31;11(3):55-61. [doi: 10.21315/eimj2019.11.3.6]
18. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Wong K, Gallant F, Szumacher E. Perceptions of Canadian radiation oncologists, radiation physicists, radiation therapists and radiation trainees about the impact of artificial intelligence in radiation oncology - national survey. J Med Imaging Radiat Sci 2021 Mar;52(1):44-48. [doi: 10.1016/j.jmir.2020.11.013] [Medline: 33323332]
21. Scheetz J, Rothschild P, McGuinness M, Hadoux X, Soyer HP, Janda M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 04;11(1):5193 [FREE Full text] [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
22. Pinto Dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
23. Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? Radiother Oncol 2018 Dec;129(3):421-426. [doi: 10.1016/j.radonc.2018.05.030] [Medline: 29907338]
24. Yüzbaşıoğlu E. Attitudes and perceptions of dental students towards artificial intelligence. J Dent Educ 2021 Jan 26;85(1):60-68. [doi: 10.1002/jdd.12385] [Medline: 32851649]
25. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]

Abbreviations

AI: artificial intelligence
ML: machine learning
OR: odds ratio
PT: physical therapist

Edited by R Kukafka; submitted 14.05.22; peer-reviewed by A Soman, M Noohu; comments to author 11.06.22; revised version received 22.06.22; accepted 28.09.22; published 20.10.22.

Please cite as:
Alsobhi M, Khan F, Chevidikunnan MF, Basuodan R, Shawli L, Neamatallah Z
Physical Therapists' Knowledge and Attitudes Regarding Artificial Intelligence Applications in Health Care and Rehabilitation: Cross-sectional Study
J Med Internet Res 2022;24(10):e39565
URL: https://www.jmir.org/2022/10/e39565
doi:10.2196/39565
PMID:36264614

Original Paper

Automatic Assessment of Intelligibility in Noise in Parkinson Disease: Validation Study

Gemma Moya-Galé[1*], PhD; Stephen J Walsh[2*], PhD; Alireza Goudarzi[3*], PhD

[1]Department of Communication Sciences & Disorders, Long Island University, Brooklyn, NY, United States

[2]Department of Mathematics and Statistics, Utah State University, Logan, UT, United States

[3]Factorize, Tokyo, Japan

[*]all authors contributed equally

Corresponding Author:
Gemma Moya-Galé, PhD
Department of Communication Sciences & Disorders
Long Island University
1 University Plaza
Brooklyn, NY, 11201
United States
Phone: 1 718 780 4125
Email: gemma.moya-gale@liu.edu

Abstract

Background: Most individuals with Parkinson disease (PD) experience a degradation in their speech intelligibility. Research on the use of automatic speech recognition (ASR) to assess intelligibility is still sparse, especially when trying to replicate communication challenges in real-life conditions (ie, noisy backgrounds). Developing technologies to automatically measure intelligibility in noise can ultimately assist patients in self-managing their voice changes due to the disease.

Objective: The goal of this study was to pilot-test and validate the use of a customized web-based app to assess speech intelligibility in noise in individuals with dysarthria associated with PD.

Methods: In total, 20 individuals with dysarthria associated with PD and 20 healthy controls (HCs) recorded a set of sentences using their phones. The Google Cloud ASR API was used to automatically transcribe the speakers' sentences. An algorithm was created to embed speakers' sentences in +6-dB signal-to-noise multitalker babble. Results from ASR performance were compared to those from 30 listeners who orthographically transcribed the same set of sentences. Data were reduced into a single event, defined as a success if the artificial intelligence (AI) system transcribed a random speaker or sentence as well or better than the average of 3 randomly chosen human listeners. These data were further analyzed by logistic regression to assess whether AI success differed by speaker group (HCs or speakers with dysarthria) or was affected by sentence length. A discriminant analysis was conducted on the human listener data and AI transcriber data independently to compare the ability of each data set to discriminate between HCs and speakers with dysarthria.

Results: The data analysis indicated a 0.8 probability (95% CI 0.65-0.91) that AI performance would be as good or better than the average human listener. AI transcriber success probability was not found to be dependent on speaker group. AI transcriber success was found to decrease with sentence length, losing an estimated 0.03 probability of transcribing as well as the average human listener for each word increase in sentence length. The AI transcriber data were found to offer the same discrimination of speakers into categories (HCs and speakers with dysarthria) as the human listener data.

Conclusions: ASR has the potential to assess intelligibility in noise in speakers with dysarthria associated with PD. Our results hold promise for the use of AI with this clinical population, although a full range of speech severity needs to be evaluated in future work, as well as the effect of different speaking tasks on ASR.

(J Med Internet Res 2022;24(10):e40567) doi:10.2196/40567

KEYWORDS

automatic speech recognition; Parkinson disease; intelligibility; dysarthria; digital health; artificial intelligence

Introduction

Parkinson disease (PD) is the second most common neurodegenerative disease, following Alzheimer disease [1]. Approximately 1 million individuals are estimated to be affected by the disease in the United States [2], and its prevalence surpasses 6 million people worldwide [3], with numbers projected to increase in the future [2]. Close to 90% of individuals with PD evidence problems with voice or speech, an impairment known as hypokinetic dysarthria, which has a latency that averages 7 years post–disease onset [4]. This motor speech disorder is characterized by hypophonia (ie, reduced loudness), monopitch, monoloudness, articulatory imprecision, reduced stress, short rushes of speech, and variable rate [5]. As a result, many individuals affected by the disease complain of intelligibility problems (ie, their ability to be understood by others) [6], especially in noisy environments (eg, when dining out at a restaurant). Additionally, the presence of background noise has been shown to negatively affect even speakers with mildly dysarthric speech [7]. Overall, these speech deficits substantially reduce speakers' social participation and overall quality of life [8], as their inability to effectively communicate with others increases their frustration and social isolation.

The application of artificial intelligence (AI) in the medical field has brought promising results to enhance communication and, ultimately, quality of life [9] in a wide range of individuals. For example, voice-assisted technology, which is used in devices such as Siri or Alexa, has become increasingly more present among individuals with a neurodegenerative disease, such as those with PD [10], and has gradually been incorporated as a potential available tool for health professionals, such as speech and language pathologists [11]. The development of automatic speech recognition (ASR) technologies has substantially advanced in the past 40 years, especially given the onset of deep learning mechanisms [12]. Most crucially, the use of ASR has been shown to be effective in estimating speakers' intelligibility deficits for different clinical populations who may present with speech impairments [13], such as those resulting from a laryngectomy [14], a cleft palate [15], or head and neck cancer [16]. Additionally, the clinical validity of ASR has also been explored in individuals with apraxia of speech and aphasia with promising results [17,18]. Project Euphonia has achieved a large-scale data set with over 1 million recordings of disordered speech, with the ultimate goal to personalize ASR models to enhance communication in individuals who experience speech and language difficulties [19,20]. Despite the great advancements that these findings represent, however, research on the application of ASR for individuals with the motor speech disorder of dysarthria has been more limited [21-23], and it has underscored the high degree of variability that characterizes dysarthric speech [13], especially with increased speech severity levels [24]. Dimauro et al [25] explored the use of ASR with 28 individuals with dysarthria associated with PD, 22 healthy older adults, and 15 healthy young controls. In their study, the speech-to-text system focused on the recognition error rates of words from different speech tasks. Although their results upheld the use of AI as a promising resource for clinical populations, it is important to note, however, that their experiment was conducted in quiet conditions, which may not reflect the real-life challenges speakers with PD face in everyday communication. More recently, Gutz et al [26] used the Google Cloud ASR API for intelligibility measurement with 52 speakers with dysarthria associated with amyotrophic lateral sclerosis and 20 healthy controls. Additionally, the authors used noise-augmented ASR to assist the AI system in discriminating between healthy speech and mildly dysarthric speech. Results from their study showed high variability and poor internal validity of machine word recognition rate, suggesting that this technology may have limited clinical applicability for this population at this time.

Our previous pilot work examined ASR performance in multitalker babble noise to measure speech intelligibility from a reading task in 5 speakers with PD and 5 healthy adults [27]. Preliminary results supported the feasibility of AI technologies to simulate real-life challenges posed by ambient noise. Our current study was aimed at expanding our previous work with speakers with dysarthria associated with PD to preliminarily validate the use of ASR in noise with this clinical population. To that end, this study reports on the development, pilot-testing, and validation of a web-based app, *Understand Me for Life* [27], to assess speech intelligibility in noise using the Google Cloud ASR API in speakers with dysarthria associated with PD. Specifically, our aims were to (1) examine how ASR compared to human transcription, the current gold standard, when determining intelligibility accuracy scores for speakers with hypokinetic dysarthria associated with PD; and (2) determine the extent to which ASR could accurately discriminate between speakers with dysarthria and healthy controls.

Methods

Ethics Approval

This study was approved by the Institutional Review Board at Long Island University, Brooklyn (21/01-002-Bkln).

Speakers

In total, 20 individuals with PD (12 women and 8 men; mean age 73.3 years; age range 62-81 years) and 20 age- and sex-matched neurologically healthy adults participated in the speech recordings for this study. Individuals with PD had to meet the following inclusion criteria: (1) having a medical diagnosis of PD, (2) having experienced changes in their voice that represented a current concern, (3) having a stable anti-Parkinsonian medication, (4) passing the Montreal Cognitive Assessment [28], and (5) being a native speaker of English. Exclusion criteria included having received intensive voice-focused treatment in the past 2 years prior to the study and having received deep brain stimulation. Neurologically healthy speakers (12 women and 8 men; mean age 70.5 years; age range 59-84 years) with no history of motor speech impairments served as controls. Table 1 presents the speakers' biographical details and clinical characteristics.

Dysarthria severity ranged from mild to moderate in these speakers and was assessed from a conversation sample by an experienced speech and language pathologist. Consensus with a second speech and language pathologist was obtained for the final dysarthria severity estimates [29].

Table 1. Speakers' biographical details and clinical characteristics.

Speaker	Age (years)	Sex	YPD[a]	Dysarthria severity	Patient's voice complaint
P1[b]	77	Female	9	Mild	Voice is softer and sounds are not as well-articulated
P2	77	Male	1	Mild-moderate	Voice is softer
P3	70	Female	6	Mild	Hoarseness
P4	72	Female	4	Mild	Less control over shaping words, changes in loudness, and occasional rapid breathing
P5	72	Female	7	Mild-moderate	Voice is much lower and softer and reduced intelligibility
P6	80	Female	8	Mild-moderate	Increased fatigue, hoarseness, and lack of clarity
P7	80	Female	8	Mild	Reduced fundamental frequency range for singing and "scratchy feeling" in throat
P8	67	Female	9	Mild-moderate	Lower pitch, hoarseness, voice is much softer, and reduced intelligibility
P9	65	Female	5	Mild	Recent coughing, softness of voice, and voice sounds rougher and softer than usual.
P10	78	Female	7	Mild	Slurring, voice is softer, and intelligibility has been affected.
P11	60	Female	8	Mild	Occasional reduction in loudness
P12	66	Male	7	Mild	Fluctuations in voice and voice is much softer
P14	73	Male	8	Mild	Occasional reduction in loudness and stuttering
P14	80	Female	7	Mild-moderate	Voice is softer
P15	73	Male	13	Mild-moderate	Voice is softer and more strained
P16	78	Male	4	Mild	Voice is softer, trouble finding words, and sometimes intelligibility is affected
P17	62	Male	13	Moderate	Voice is very soft, problems with intelligibility, and fast speaking rate
P18	81	Male	8	Mild-moderate	Voice is softer, breathiness, and have to clear throat more often
P19	80	Female	8	Mild	Voice is softer
P20	76	Male	7	Moderate	Soft voice and hoarseness
HC1[c]	68	Female	N/A[d]	N/A	N/A
HC2	71	Male	N/A	N/A	N/A
HC3	64	Female	N/A	N/A	N/A
HC4	67	Male	N/A	N/A	N/A
HC5	72	Female	N/A	N/A	N/A
HC6	77	Female	N/A	N/A	N/A
HC7	72	Male	N/A	N/A	N/A
HC8	71	Male	N/A	N/A	N/A
HC9	67	Female	N/A	N/A	N/A
HC10	78	Male	N/A	N/A	N/A
HC11	59	Female	N/A	N/A	N/A
HC12	61	Male	N/A	N/A	N/A
HC13	75	Female	N/A	N/A	N/A
HC14	66	Female	N/A	N/A	N/A
HC15	63	Female	N/A	N/A	N/A
HC16	63	Male	N/A	N/A	N/A
HC17	84	Female	N/A	N/A	N/A
HC18	84	Male	N/A	N/A	N/A

Speaker	Age (years)	Sex	YPD[a]	Dysarthria severity	Patient's voice complaint
HC19	65	Female	N/A	N/A	N/A
HC20	83	Female	N/A	N/A	N/A

[a]YPD: years postdiagnosis.

[b]P: patient (speaker with dysarthria associated with Parkinson disease).

[c]HC: healthy control.

[d]N/A: not applicable.

Speech Stimuli and Recording Procedures

A set of 100 grammatically and semantically correct sentences was created for this study. Sentences differed in length, from 5 to 9 words (eg, "Take care of my house while I am away"), and contained high frequency words in the English language (The English Lexicon Project) [30]. The data set was then divided into 4 different blocks of 25 randomized sentences each, with blocks having an equal number of sentences from each sentence length. Each speaker was randomized to 1 block of stimuli for speech recordings, so that each block was read by 10 different speakers. Recordings were self-paced and conducted in a quiet room in the speakers' homes using a customized web-based app, *Understand Me for Life* [27], that the speakers could access from their mobile phones. The first author met with speakers over the Zoom videoconferencing platform (Zoom Video Communications) to explain the recording procedure and address any potential questions. Careful directions were provided to ensure a constant 8-cm (3.15 inches) mouth-to-microphone distance [31,32]. Given the possibility of PD-related motor impairments hindering adequate recordings (eg, tremors), care partners were recruited to assist speakers when necessary. Speakers were allowed to rerecord a sentence in cases of extraneous noise in the background. A brief familiarization phase was provided at the beginning of the recording session so that speakers could practice using the interface. Feedback from speakers was obtained for later app optimization.

For each recorded sentence, the app automatically embedded the speakers' voice signal into +6-dB signal-to-noise multitalker babble noise [33] to provide an intelligibility score, defined as the percentage of words accurately understood by the ASR system. Automatic feedback on performance was provided at the end of the recording session and not after each sentence to avoid any potential priming effects that could influence sentence production on subsequent items [34].

Multitalker Babble Noise

Multitalker babble is thought to be the most common type of environmental noise experienced by listeners [35], which, therefore, makes it more ecologically valid in speech perception experiments. For this study, 10-second sample recordings from National Public Radio were used. Audio files were manually checked to control for sudden changes in the speech signal (eg, increase in vocal intensity). Prolonged silences (ie, over 500 ms) were trimmed, followed by the equalization of the audio spectrum in a moving window. An equal number of male and female speakers was implemented in the creation of background noise [36]. The equalized audios were finally combined to render 10-talker babble [33].

Listeners

In total, 30 neurologically healthy adults (25 women and 5 men; mean age 23.1 years; age range 18-31 years) participated as listeners in the study. Listeners were recruited via flyers and word of mouth across the New York City area. Inclusion criteria for participation required listeners to be native speakers of English; have no history of speech, language, or communication impairment; have no prior experience with motor speech disorders; and pass a bilateral pure-tone hearing screening at 25-dB hearing level at 500, 1000, 2000, and 4000 Hz [37]. Listeners were paid US $20 for their participation in the study.

Human Transcription

Listeners completed the intelligibility assessment task free field (ie, without headphones) in a quiet space at the Long Island University campus, in Brooklyn, New York. The task was accessible through the *Understand Me for Life* portal on a MacBook Pro laptop (Apple Inc). Listeners maintained a distance of 85 cm from the loudspeakers (Logitech Z150), and the loudspeakers were placed 31 cm from each other. Listener-to-loudspeaker distance represented the typical distance between conversational partners [38]. The task took approximately 30-40 minutes to complete.

A brief familiarization phase was presented before the start of the experiment and contained 3 sentences produced by a neurologically healthy adult male speaker. Listeners were instructed to write down word by word what they heard and not worry about punctuation marks. Each listener was randomly assigned to 1 speaker per block, with block presentation being random across listeners. Therefore, each listener heard a total of 4 speakers and 100 sentences. Sentences were presented in multitalker babble, hence replicating the AI condition. To avoid abrupt onsets and offsets of stimuli, 400 ms of noise were inserted at the beginning of each sentence, and each sentence was followed by 50 ms of babble noise [39]. To obtain an average score for subsequent transcription accuracy calculations, each speaker was assigned to 3 listeners. None of the listeners required a break during the completion of this task.

Data Analysis

Automatic Intelligibility Assessment

Automatic intelligibility assessment (AIA) was conducted using the Google Cloud ASR API, a speech-to-text AI system with documented low word error rate for individuals with healthy speech that is thought to be the best platform to handle dysarthric speech, although software performance is still dependent on speech severity, with high word error rates in cases of more severely affected speech [40].

For a given produced utterance (*S*) and the corresponding target sentence (*T*), stimuli were suitably padded with whitespace to ensure that both *S* and *T* were of equal length (*L*). Each word in *S* was codified with w_s and each word in *T* with w_t, where *s* and *t* were numbers from 0 to $L - 1$. Accuracy was calculated by the formula as follows:

where $\sigma(w_s,w_t) = 1$ if $w_s = w_t$, and 0 otherwise. This step was implemented to avoid providing a score to words that appeared in both *S* and *T* but were out of order [27].

Manual Intelligibility Assessment

Transcription accuracy scores were calculated as the percentage of words correctly transcribed. Orthographic transcriptions are considered the most objective measure to assess intelligibility in dysarthria [33]. Listeners' orthographic transcripts had to match the target to be accepted as correct [32,41]. Obvious spelling errors or errors involving homonyms did not impact calculation scores and were assessed as correct responses. Omissions or additions of morphemes (eg, flower for flowers) were coded as an error.

Statistical Analysis

The goal of the first phase of statistical analysis was to assess the degree to which the AIA could score as well or better than the average human transcriber (ie, listener). As described above, 3 listeners orthographically transcribed sentences from the same speakers, and their data were condensed into a *percentage accuracy* measure for each sentence, which summarized the percentage of words the human listener correctly transcribed. For each question, the average percentage accuracy, denoted as $\hat{a}_{ij,\ human\ avg}$, was computed for each sentence *j* within each speaker *i* to reduce intralistener variability. The AIA system also received a percentage accuracy measure for each sentence or speaker, which we denoted as $\hat{a}_{ij,\ AIA}$. The success of the AIA system was defined as follows:

The AIA system was considered to give a successful transcription if its percentage accuracy score was at least as good as the average of the human listeners' accuracies for sentence *j* within each speaker *i*. The data were then condensed up to the speaker level by computing the proportion of successes of the AIA system over the $j = 1, \ldots, 25$ sentences read by speaker *i* as follows:

This procedure provided an estimate of the probability of success of the AIA system transcription for randomly selected speakers. Standard binomial statistics were used to quantify uncertainty in this analysis and present the results with appropriate statistical summaries and CIs. We investigated whether data provided evidence that the AIA transcriber success differed whether the system was transcribing a healthy control (HC) or a speaker with dysarthria associated with PD and whether sentence length had an effect on AIA success, via a logistic regression analysis.

The goal of the second phase of statistical analysis was to compare the ability of the resulting AIA transcription data summaries to discriminate between healthy controls and speakers with dysarthria. To investigate this goal, we applied linear discriminant analysis to identify optimal discrimination thresholds for both the listener transcriptions and the AIA transcriptions and summarized the discrimination ability of each via typical confusion matrices and correct percentage classification summaries. All statistical analyses were conducted in R statistical software (version 4.1.1; R Foundation for Statistical Computing) [42] and a discriminant and classification analysis was conducted via the *lda* function in the *MASS* package [43].

Intralistener reliability was assessed via percentage agreement on several (approximately 10) duplicate speaker sentences. Interlistener reliability was controlled for in this assessment by condensing each of the 3 listeners' percentage accuracy measures for each speaker or sentence into the average.

Results

A summary of intrarater reliability is shown in Figure 1. The average percentage agreement of repeated responses of this study's listeners was 80%.

The success summaries of the AIA transcriber at the speaker level are presented in Figure 2. The figure shows estimates of the probability of success for each speaker (ordered by score) with a 95% CI. The mean probability of success is indicated by the red horizontal line. The figure illustrates that the expected success probability of the AIA transcriber for a randomly selected speaker was approximately 0.8 (95% CI 0.65-0.91), with the AIA system scoring 80% of target sentences as well or better than the human transcribers for half (22/40, 55%) of the study's speakers. The success probability estimates stratified by speaker group (HC or speaker with dysarthria) are shown in Figure 3. The figure suggests that the AIA transcriber had a slightly more difficult time accurately transcribing the sentences read by speakers with dysarthria, with a slight decline in the estimate of probability of success for speakers #14, #18, and #19.

Figure 1. Distribution of intrarater percentage agreement across the 30 listeners.

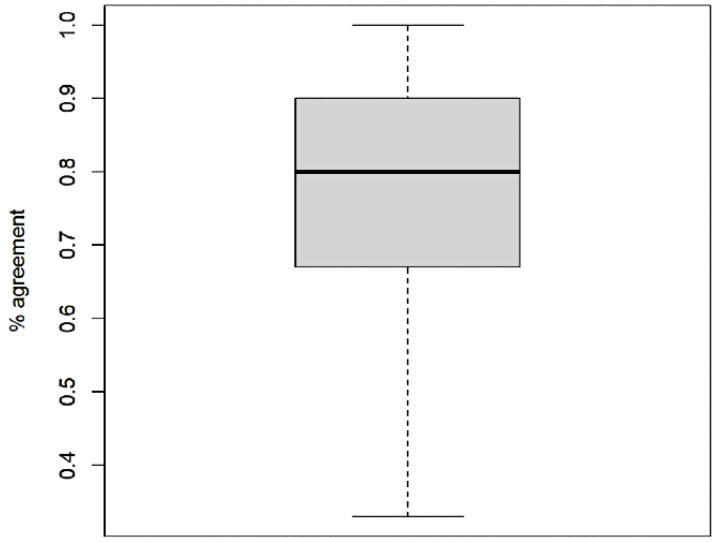


Figure 2. Estimates of the probability that the automatic intelligibility assessment transcriber will be as accurate as human transcribers for each speaker. The vertical bands are 95% CIs on the estimate of probability of success. Black dotted line=0.5 and red dotted line=median AI probability of success. AI: artificial intelligence; C: control; P: patient with dysarthria.

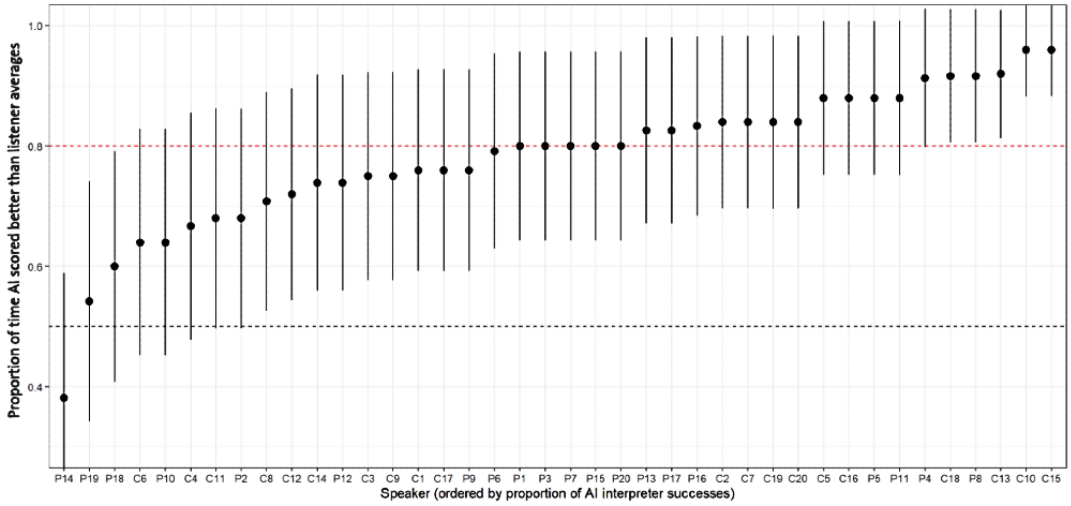
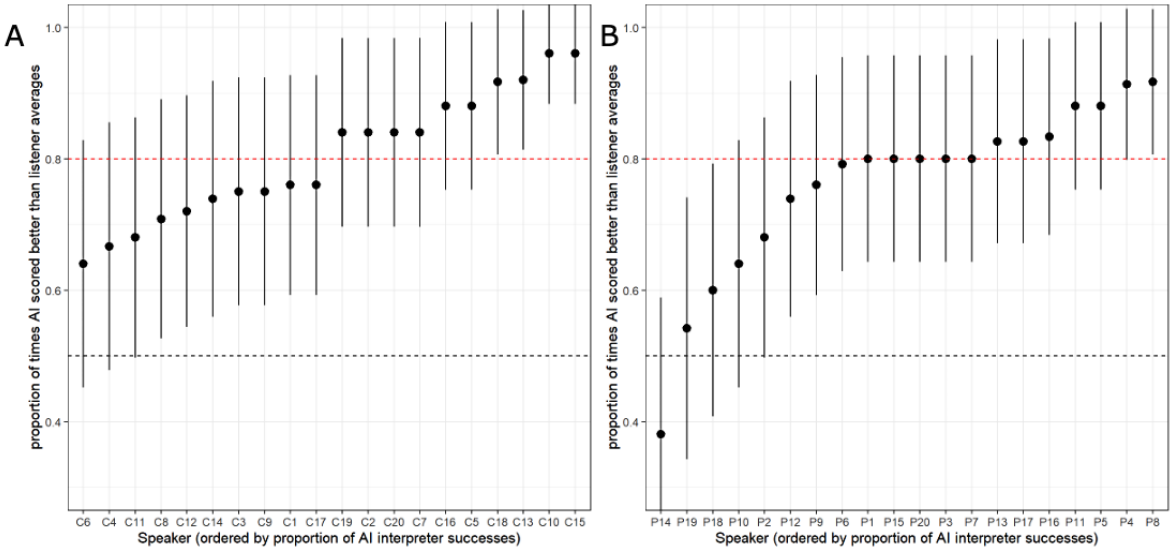


Figure 3. Estimates of the probability that the automatic intelligibility assessment transcriber will be as accurate as human transcribers for each speaker: (A) healthy controls and (B) speakers with dysarthria. AI: artificial intelligence; C: control; P: patient with dysarthria.



We further analyzed these data via a logistic regression model. The response was the (logit) probability of AI success and the predictors were speaker group (HC or speakers with dysarthria) and sentence type. Speaker-to-speaker variance was controlled for by including speaker as a random effect. The fitted model estimates are presented in Table 2. The advantage of this approach is that each row provides a significance test for each term *provided we have controlled for the effects of the other terms*. In this regard, after controlling for speaker and sentence length, we see that these data provide weak evidence that AI success differs significantly by speaker group (ie, between HC and speakers with dysarthria; P=.23). Further, sentence length was found to have a significant negative impact on AI success (P<.001). The results are represented in an effects plot in Figure 4. The left panel illustrates that an estimate of the probability of AI success for speakers with dysarthria is 0.78, but this value is not significantly different from the estimate of the probability of AI success for HCs (0.82; P=.23). The right panel illustrates an estimated dependence of the probability of AI success on sentence length, with each increase in sentence length decreasing AI success probability by an estimated 0.03.

Percentage accuracy distributions by transcriber (human or AIA system) and speaker group are presented in Figure 5. The box plots in Figure 5 indicate that the median accuracy score for speakers with dysarthria was farther from the median accuracy score for healthy controls as compared to the distance between the 2 medians for the human transcriber data. This finding suggests that the AIA system data may offer better discrimination and classification ability for speaker group.

Confusion matrices recording the classification rates of discriminants based on human transcription data and AIA system data are presented in Table 3.

Table 2. Fitted logistic regression model coefficients.

Effect	Estimate	SE	*z* value	*P* value
Intercept	3.14414	0.44774	7.022	<.001
Speaker group	–0.25525	0.21156	–1.207	.23
Sentence length	–0.23658	0.05763	–4.105	.001

Figure 4. Estimated effects and CIs from the logistic regression of probability of AI success as a function of (A) speaker group, (B) sentence length, and speaker random effect. AI: artificial intelligence; HC: healthy controls.

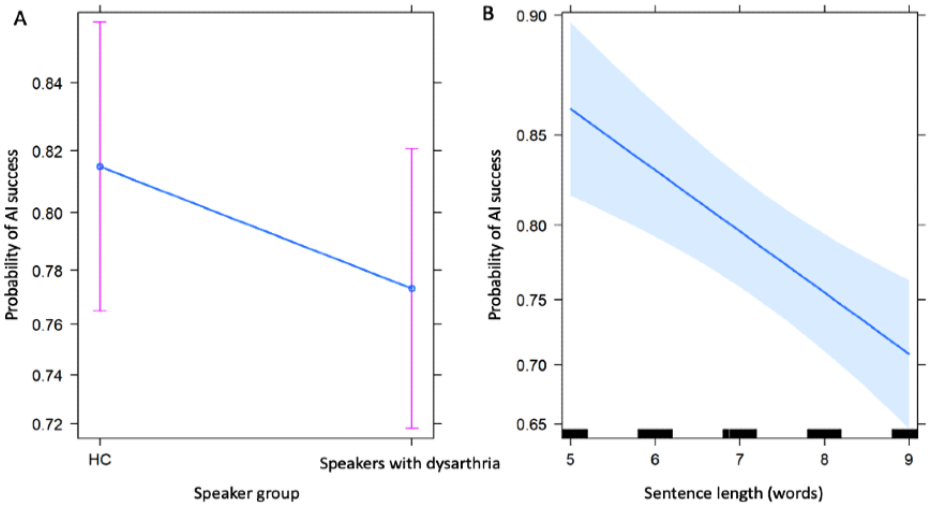


Figure 5. Box plots of the estimates of AIA system success by speaker category and transcriber: (A) human listener and (B) AIA system. AI: artificial intelligence; AIA: automatic intelligibility assessment; HC: healthy controls.

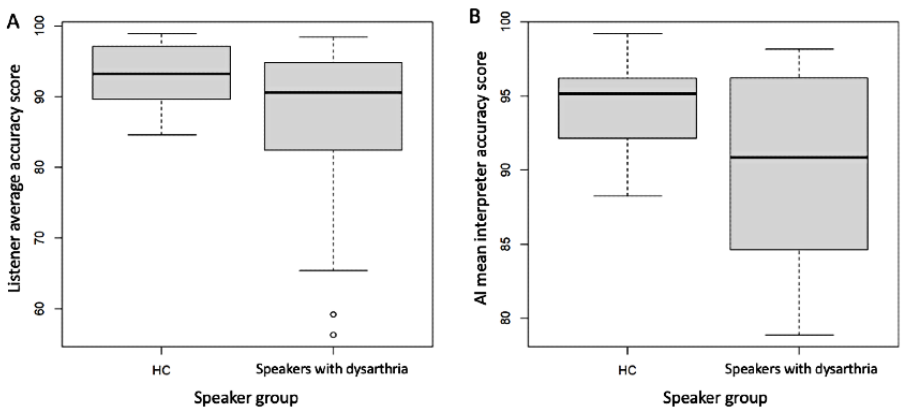


Table 3. Classification summary of the speakers based on linear discriminants fit to the human transcription data and automatic intelligibility assessment system data.

True group	Classified group via discriminant			
	Discriminant from human listener average data (overall predictive accuracy: 0.6)		Discriminant from artificial intelligence data (overall predictive accuracy: 0.675)	
	HC[a]	PD[b]	HC	PD
HC	15	5	15	5
PD	11	9	8	12

[a]HC: healthy control.

[b]PD: Parkinson disease.

Discussion

Principal Findings

This study aimed to develop, pilot-test, and validate the use of a web-based app, *Understand Me for Life*, to automatically measure speech intelligibility in noise in speakers with hypokinetic dysarthria associated with PD. Additionally, a secondary objective of the study was to determine whether ASR could discriminate between the speech of healthy controls and that of speakers with dysarthria.

Literature on ASR performance on clinical populations, especially those with motor speech disorders, is still sparse. To validate the use of speech-to-text technology to determine intelligibility accuracy scores for speakers with dysarthria, ASR performance was benchmarked relative to that of human transcribers [19]. Results showed that the ASR system had an 80% chance of performing as well as or better than a human transcriber on any random speaker. The potential capacity of ASR to outperform human listeners has been shown in recent studies [19], although further work is required with longer utterances and different speech tasks, as summarized in the limitations section below. Our findings also echo those reported with other clinical populations, such as those with a diagnosis of apraxia of speech and aphasia [17,18]. Additionally, our data provided no evidence that the mean probability of ASR success differed between the 2 groups of speakers, either a speaker with dysarthria or a healthy control. Thus, the success of the speech-to-text system did not depend on whether the speaker was neurologically healthy or presented with hypokinetic dysarthria associated with PD. It is important to acknowledge, however, that our speakers did not evidence dysarthria across all severity ranges; this limitation will be addressed in future work. Sentence length did influence ASR, with a decrease in accuracy observed for longer sentences, which was an expected result and is in agreement with prior literature [19,26].

The second aim of the study was to determine whether ASR could accurately discriminate between speakers with dysarthria and healthy controls. Results showed that both the human and the AIA system data provided the same classification rates for healthy controls (15/20, 75% correctly classified and 5/20, 25% incorrectly classified as speakers with dysarthria), hence evidencing equal specificity (ie, 75%). The AIA system data, however, yielded a slightly better classification success for speakers with dysarthria (12/20, 60% correct PD classifications compared to the human transcription data that only yielded 9/20, 45% correct PD classifications), which suggests stronger sensitivity than the one obtained for human transcribers (ie, 60% vs 45%). In traditional studies using human listeners, performance on intelligibility assessments has not shown significant differences between speakers with mild dysarthria secondary to PD and healthy controls [33], hence suggesting that group classification based on intelligibility scores may depend on speech severity. In our study, AI correctly classified 12 speakers with dysarthria (out of 20), a result that could be explained by the severity levels of our sample ranging from mild to mild-to-moderate only.

Limitations and Future Work

The study's limitations warrant future work in this research area. It should be noted that our sample of speakers with dysarthria did not include those with more severe speech deficits. Therefore, these results offer a preliminarily promising, albeit not conclusive, clinical tool for measuring intelligibility in individuals with dysarthria associated with PD. Nevertheless, ASR performance with a more diverse speech severity range in speakers with dysarthria associated with PD should be explored. It is likely that increased speech severity in individuals with PD would impact ASR, as this increase was also found in speakers with dysarthria associated with amyotrophic lateral sclerosis [26]. An additional limitation from this study is that the speech stimuli were derived from read sentences rather than from conversational speech. Although sentences rendered a higher level of predictability and, thus, control, conversational speech would have greater ecological validity. Finally, we should also acknowledge that previously reported studies used different ASR methodology compared to this study and that, as discussed in Jacks et al [18], ASR technology is in constant and rapid evolution, rendering any results on ASR in need of systematic reevaluation for the proper and valid use of ASR-assisted clinical tools.

Our ongoing work is motivated by the concept of self-management, which, in the context of a chronic illness such as PD, has become increasingly relevant. Self-management relates to the patient's ability to identify a given behavior (eg, voice changes) and react or problem-solve in accordance with such observation [44]. Having the knowledge on how to respond to the worsening of disease symptoms and when to seek medical advice has been shown to be crucial contributors to patients' well-being [45]. The implementation of ASR in speech intelligibility assessment, therefore, can potentially serve to

establish preventative measures before the onset of speech and intelligibility degradation and control measures (eg, referral to a speech therapist) if speech deficits already exist.

Conclusions

This study validated the use of ASR to measure intelligibility in real-life settings (ie, using background noise) in speakers with mild-to-moderate dysarthria associated with PD. Therefore, our preliminary data show that ASR has the potential to assess intelligibility in noise in this clinical population. Results hold promise for the use of AI as a future clinical tool to assist patients and speech and language therapists alike, although the full range of speech severity needs to be evaluated in future work, as well as the effect of different speaking tasks on ASR.

Acknowledgments

We wholeheartedly thank the participants in this study, their care partners, as well as our research assistant, Robert Seefeldt, for his priceless help across the different stages of the project. This project was funded by the Michael J. Fox Foundation for Parkinson's Research (grant 001236; awarded to GM-G, the principal investigator).

Conflicts of Interest

None declared.

References

1. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007 Jan 30;68(5):384-386. [doi: 10.1212/01.wnl.0000247740.47667.03] [Medline: 17082464]
2. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, Parkinson's Foundation P4 Group. Prevalence of Parkinson's disease across North America. NPJ Parkinsons Dis 2018 Jul 10;4:21 [FREE Full text] [doi: 10.1038/s41531-018-0058-0] [Medline: 30003140]
3. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. J Parkinsons Dis 2018 Dec 18;8(s1):S3-S8 [FREE Full text] [doi: 10.3233/JPD-181474] [Medline: 30584159]
4. Müller J, Wenning GK, Verny M, McKee A, Chaudhuri KR, Jellinger K, et al. Progression of dysarthria and dysphagia in postmortem-confirmed Parkinsonian disorders. Arch Neurol 2001 Feb 01;58(2):259-264. [doi: 10.1001/archneur.58.2.259] [Medline: 11176964]
5. Duffy JR. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. 4th ed. Amsterdam, the Netherlands: Elsevier Mosby; 2020.
6. Moya-Galé G, Rossi A, States RA. A community-based program for exercise and social participation for individuals with Parkinson's disease: a multidisciplinary model. Perspect ASHA SIGs 2020 Oct 23;5(5):1290-1296. [doi: 10.1044/2020_persp-20-00031]
7. Chiu Y, Forrest K. The impact of lexical characteristics and noise on intelligibility of Parkinsonian speech. J Speech Lang Hear Res 2018 Apr 17;61(4):837-846. [doi: 10.1044/2017_JSLHR-S-17-0205] [Medline: 29587306]
8. McAuliffe MJ, Baylor CR, Yorkston KM. Variables associated with communicative participation in Parkinson's disease and its relationship to measures of health-related quality-of-life. Int J Speech Lang Pathol 2017 Aug 27;19(4):407-417 [FREE Full text] [doi: 10.1080/17549507.2016.1193900] [Medline: 27347713]
9. Derosier R, Farber RS. Speech recognition software as an assistive device: a pilot study of user satisfaction and psychosocial impact. Work 2005;25(2):125-134. [Medline: 16131742]
10. Duffy O, Synnott J, McNaney R, Brito Zambrano P, Kernohan WG. Attitudes toward the use of voice-assisted technologies among people with Parkinson disease: findings from a web-based survey. JMIR Rehabil Assist Technol 2021 Mar 11;8(1):e23006 [FREE Full text] [doi: 10.2196/23006] [Medline: 33704072]
11. Kulkarni P, Duffy O, Synnott J, Kernohan WG, McNaney R. Speech and language practitioners' experiences of commercially available voice-assisted technology: web-based survey study. JMIR Rehabil Assist Technol 2022 Jan 05;9(1):e29249 [FREE Full text] [doi: 10.2196/29249] [Medline: 34989694]
12. Kodish-Wachs J, Agassi E, Kenny PIII, Overhage JM. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. AMIA Annu Symp Proc 2018 Dec 05;2018:683-689 [FREE Full text] [Medline: 30815110]
13. Tu M, Wisler A, Berisha V, Liss JM. The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. J Acoust Soc Am 2016 Nov;140(5):EL416-EL422 [FREE Full text] [doi: 10.1121/1.4967208] [Medline: 27908075]
14. Schuster M, Haderlein T, Nöth E, Lohscheller J, Eysholdt U, Rosanowski F. Intelligibility of laryngectomees' substitute speech: automatic speech recognition and subjective rating. Eur Arch Otorhinolaryngol 2006 Feb 7;263(2):188-193. [doi: 10.1007/s00405-005-0974-6] [Medline: 16001246]

15. Maier A, Haderlein T, Stelzle F, Nöth E, Nkenke E, Rosanowski F, et al. Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer. EURASIP J Audio Speech Music Process 2010 Dec 01;2010(1):1-7 [FREE Full text]
16. Maier A, Nöth E, Nkenke E, Schuster M. Automatic assessment of children's speech with cleft lip and palate. 2006 Oct Presented at: IS-LTC 2006: 5th Slovenian and 1st International Conference on Language Technologies; October 9-10, 2006; Ljubljana, Slovenia URL: http://nl.ijs.si/is-ltc06/proc/06_Maier.pdf
17. Ballard KJ, Etter NM, Shen S, Monroe P, Tien Tan C. Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for apraxia of speech plus aphasia. Am J Speech Lang Pathol 2019 Jul 15;28(2S):818-834. [doi: 10.1044/2018_AJSLP-MSC18-18-0109] [Medline: 31306595]
18. Jacks A, Haley KL, Bishop G, Harmon TG. Automated speech recognition in adult stroke survivors: comparing human and computer transcriptions. Folia Phoniatr Logop 2019 May 22;71(5-6):286-296 [FREE Full text] [doi: 10.1159/000499156] [Medline: 31117105]
19. Green JR, MacDonald RL, Jiang PP, Cattiau J, Heywood R, Cave R, et al. Automatic speech recognition of disordered speech: personalized models now outperforming human listeners on short phrases. 2021 Presented at: Interspeech 2021; August 30 to September 3, 2021; Brno, Czechia p. 4778-4782. [doi: 10.21437/interspeech.2021-1384]
20. MacDonald RL, Jiang PP, Cattiau J, Heywood R, Cave R, Seaver K, et al. Disordered speech data collection: lessons learned at 1 million utterances from Project Euphonia. 2021 Presented at: Interspeech 2021; August 30 to September 3, 2021; Brno, Czechia p. 4833-4837. [doi: 10.21437/interspeech.2021-697]
21. Christensen H, Cunningham S, Fox C, Green P, Hain T. A comparative study of adaptive, automatic recognition of disordered speech. 2012 Presented at: Interspeech 2012: 13th Annual Conference of the International Speech Communication Association (ISCA); September 9-13, 2012; Portland, OR p. 1776-1779. [doi: 10.21437/interspeech.2012-484]
22. Sharma HV, Hasegawa-Johnson M. Acoustic model adaptation using in-domain background models for dysarthric speech recognition. Comput Speech Lang 2013 Sep;27(6):1147-1162. [doi: 10.1016/j.csl.2012.10.002]
23. Vásquez-Correa JC, Orozco-Arroyave JR, Bocklet T, Nöth E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. J Commun Disord 2018 Nov;76:21-36. [doi: 10.1016/j.jcomdis.2018.08.002] [Medline: 30149241]
24. Le D, Licata K, Mower Provost E. Automatic quantitative analysis of spontaneous aphasic speech. Speech Commun 2018 Jun;100:1-12. [doi: 10.1016/j.specom.2018.04.001]
25. Dimauro G, Di Nicola V, Bevilacqua V, Caivano D, Girardi F. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. IEEE Access 2017 Oct 17;5:22199-22208. [doi: 10.1109/access.2017.2762475]
26. Gutz SE, Stipancic KL, Yunusova Y, Berry JD, Green JR. Validity of off-the-shelf automatic speech recognition for assessing speech intelligibility and speech severity in speakers with amyotrophic lateral sclerosis. J Speech Lang Hear Res 2022 Jun 08;65(6):2128-2143. [doi: 10.1044/2022_JSLHR-21-00589] [Medline: 35623334]
27. Goudarzi A, Moya-Galé G. Automatic speech recognition in noise for Parkinson's disease: a pilot study. Front Artif Intell 2021 Dec 22;4:809321 [FREE Full text] [doi: 10.3389/frai.2021.809321] [Medline: 35005616]
28. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
29. Fletcher AR, McAuliffe MJ, Lansford KL, Liss JM. Assessing vowel centralization in dysarthria: a comparison of methods. J Speech Lang Hear Res 2017 Feb 01;60(2):341-354 [FREE Full text] [doi: 10.1044/2016_JSLHR-S-15-0355] [Medline: 28124069]
30. Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, et al. The English Lexicon Project. Behav Res Methods 2007 Aug;39(3):445-459. [doi: 10.3758/bf03193014] [Medline: 17958156]
31. Levy ES, Moya-Galé G, Chang YM, Campanelli L, MacLeod AAN, Escorial S, et al. Effects of speech cues in French-speaking children with dysarthria. Int J Lang Commun Disord 2020 May 20;55(3):401-416. [doi: 10.1111/1460-6984.12526] [Medline: 32077196]
32. Moya-Galé G, Keller B, Escorial S, Levy ES. Speech treatment effects on narrative intelligibility in French-speaking children with dysarthria. J Speech Lang Hear Res 2021 Jun 18;64(6S):2154-2168. [doi: 10.1044/2020_JSLHR-20-00258] [Medline: 33719503]
33. Chiu YF, Neel A. Predicting intelligibility deficits in Parkinson's disease with perceptual speech ratings. J Speech Lang Hear Res 2020 Feb 26;63(2):433-443. [doi: 10.1044/2019_JSLHR-19-00134] [Medline: 32097080]
34. Maas E, Robin DA, Austermann Hula SN, Freedman SE, Wulf G, Ballard KJ, et al. Principles of motor learning in treatment of motor speech disorders. Am J Speech Lang Pathol 2008 Aug;17(3):277-298. [doi: 10.1044/1058-0360(2008/025)] [Medline: 18663111]
35. Fontan L, Tardieu J, Gaillard P, Woisard V, Ruiz R. Relationship between speech intelligibility and speech comprehension in babble noise. J Speech Lang Hear Res 2015 Jun;58(3):977-986. [doi: 10.1044/2015_JSLHR-H-13-0335] [Medline: 25809922]

36. Moya-Galé G, Goudarzi A, Bayés À, McAuliffe M, Bulté B, Levy ES. The effects of intensive speech treatment on conversational intelligibility in Spanish speakers with Parkinson's disease. Am J Speech Lang Pathol 2018 Feb 06;27(1):154-165. [doi: 10.1044/2017_AJSLP-17-0032] [Medline: 29351354]
37. ANSI S3.6-2004: specifications for audiometers. American National Standards Institute. 2004 May 13. URL: https://webstore.ansi.org/Standards/ASA/ansis32004 [accessed 2022-10-03]
38. Hall ET. The hidden dimension. Leonardo 1973;6(1):94. [doi: 10.2307/1572461]
39. Levy ES, Moya-Galé G, Chang YHM, Freeman K, Forrest K, Brin MF, et al. The effects of intensive speech treatment on intelligibility in Parkinson's disease: a randomised controlled trial. EClinicalMedicine 2020 Jul;24:100429 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100429] [Medline: 32639484]
40. De Russis L, Corno F. On the impact of dysarthric speech on contemporary ASR cloud platforms. J Reliable Intell Environ 2019 Jul 6;5(3):163-172. [doi: 10.1007/s40860-019-00085-y]
41. Cannito MP, Suiter DM, Beverly D, Chorna L, Wolf T, Pfeiffer RM. Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. J Voice 2012 Mar;26(2):214-219. [doi: 10.1016/j.jvoice.2011.08.014] [Medline: 22209057]
42. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-10-03]
43. Venables WN, Ripley BD. Modern Applied Statistics with S (MASS). 4th ed. New York, NY: Springer; 2002.
44. Lorig K. Self-management of chronic illness: a model for the future. Generations 1993 Oct 03;17(3):11-14 [FREE Full text]
45. Hayes C. Identifying important issues for people with Parkinson's disease. Br J Nurs 2002 Jan 17;11(2):91-97. [doi: 10.12968/bjon.2002.11.2.9309] [Medline: 11823736]

Abbreviations

AI: artificial intelligence
AIA: automatic intelligibility assessment
ASR: automatic speech recognition
HC: healthy control
PD: Parkinson disease

Edited by R Kukafka; submitted 27.06.22; peer-reviewed by G Klein, J Delgado Hernández, M Balaguer; comments to author 23.08.22; revised version received 05.09.22; accepted 16.09.22; published 20.10.22.

Please cite as:
Moya-Galé G, Walsh SJ, Goudarzi A
Automatic Assessment of Intelligibility in Noise in Parkinson Disease: Validation Study
J Med Internet Res 2022;24(10):e40567
URL: https://www.jmir.org/2022/10/e40567
doi:10.2196/40567
PMID:36264608

Review

Conversational Agents in Health Care: Scoping Review of Their Behavior Change Techniques and Underpinning Theory

Laura Martinengo[1], MD, PhD; Ahmad Ishqi Jabir[1], MSc; Westin Wei Tin Goh[1], MBBS; Nicholas Yong Wai Lo[1], MBBS; Moon-Ho Ringo Ho[2], PhD; Tobias Kowatsch[3,4,5,6], PhD; Rifat Atun[7,8,9,10], MBBS, MBA, DIC; Susan Michie[11], BA, MPhil, DPhil; Lorainne Tudor Car[1,12], MD, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]School of Social Sciences, Nanyang Technological University Singapore, Singapore, Singapore
[3]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[4]School of Medicine, University of St.Gallen, St. Gallen, Switzerland
[5]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[6]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise, Singapore-ETH Centre, Singapore, Singapore
[7]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Cambridge, MA, United States
[8]Department of Global Health and Social Medicine, Harvard Medical School, Harvard University, Cambridge, MA, United States
[9]Health Systems Innovation Lab, Harvard T.H. Chan School of Public Health, Harvard University, Cambridge, MA, United States
[10]Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Harvard University, Cambridge, MA, United States
[11]UCL Centre for Behaviour Change, University College London, London, United Kingdom
[12]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MD, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Clinical Sciences Building Level 18
Singapore, 308232
Singapore
Phone: 65 69047142
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: Conversational agents (CAs) are increasingly used in health care to deliver behavior change interventions. Their evaluation often includes categorizing the behavior change techniques (BCTs) using a classification system of which the BCT Taxonomy v1 (BCTTv1) is one of the most common. Previous studies have presented descriptive summaries of behavior change interventions delivered by CAs, but no in-depth study reporting the use of BCTs in these interventions has been published to date.

Objective: This review aims to describe behavior change interventions delivered by CAs and to identify the BCTs and theories guiding their design.

Methods: We searched PubMed, Embase, Cochrane's Central Register of Controlled Trials, and the first 10 pages of Google and Google Scholar in April 2021. We included primary, experimental studies evaluating a behavior change intervention delivered by a CA. BCTs coding followed the BCTTv1. Two independent reviewers selected the studies and extracted the data. Descriptive analysis and frequent itemset mining to identify BCT clusters were performed.

Results: We included 47 studies reporting on mental health (n=19, 40%), chronic disorders (n=14, 30%), and lifestyle change (n=14, 30%) interventions. There were 20/47 embodied CAs (43%) and 27/47 CAs (57%) represented a female character. Most CAs were rule based (34/47, 72%). Experimental interventions included 63 BCTs, (mean 9 BCTs; range 2-21 BCTs), while comparisons included 32 BCTs (mean 2 BCTs; range 2-17 BCTs). Most interventions included BCTs 4.1 "Instruction on how to perform a behavior" (34/47, 72%), 3.3 "Social support" (emotional; 27/47, 57%), and 1.2 "Problem solving" (24/47, 51%). A total of 12/47 studies (26%) were informed by a behavior change theory, mainly the Transtheoretical Model and the Social Cognitive Theory. Studies using the same behavior change theory included different BCTs.

Conclusions: There is a need for the more explicit use of behavior change theories and improved reporting of BCTs in CA interventions to enhance the analysis of intervention effectiveness and improve the reproducibility of research.

(J Med Internet Res 2022;24(10):e39243) doi:10.2196/39243

KEYWORDS

behavior change; behavior change techniques; conversational agent; chatbot; mHealth

Introduction

Conversational agents (CAs), or chatbots, are computer programs that simulate conversations with humans [1]. Although the first CAs were developed in the mid-1960s, it was not until the early 2000s that their availability and popularity markedly increased [2]. CAs can be used to automate a variety of tasks, such as the provision of news or weather forecasts and the facilitation of web-based shopping [3]. CAs may be deployed as stand-alone apps or websites, integrated into multifunctional apps, or included in messaging apps such as Telegram, Facebook Messenger, and Slack [2]. They may use text or voice-assisted interfaces or may include an embodied agent using virtual characters to simulate both verbal and nonverbal aspects of human communication [4]. CAs can be further classified as simple rule-based agents or smart, artificial intelligence (AI)–based agents using natural language processing or machine learning to generate the responses [2].

Following the trends in other industries, health care has seen increasing adoption of CAs in recent years [1]. Health care CAs are versatile tools able to cater to several health needs, such as providing timely information [5], supporting mental health disorder management [6,7], assisting with triage in clinical settings [8,9], supporting chronic disease self-management, or delivering lifestyle change interventions, such as physical activity [10] and dietary changes, that increasingly incorporate elements of behavior change in the intervention design. In general, health care CAs appear to be effective in improving individuals' outcomes [11,12] and are acceptable to users, who often describe them as friendly and trustworthy.

Increasingly, health care CAs are used to deliver behavior change interventions, defined as complex interventions, comprising an interplay of 1 or several heterogeneous behavior change techniques (BCTs) [13]. BCTs are "observable and replicable components designed to change behavior" [13]. BCTs are considered the smallest active ingredient in an intervention, and can be used alone or in combination with other BCTs [13]. Adequate categorization of the BCTs included in an intervention allows for more efficient coding, leading to easier replication when designing similar interventions [13]. Several methods to classify BCTs have been developed, of which the Behavior Change Technique Taxonomy version 1 (BCTTv1) [14] is the most established and commonly used.

Several reviews have synthesized the evidence about behavior change interventions delivered by digital health tools and CAs, such as a systematic review reporting on the use of BCTs in effective digital diabetes prevention interventions [15], a mapping review offering a description of the current uses of CAs for behavior change [16], and a scoping review describing the use of embodied CAs to support healthy lifestyle [17]. These reviews presented descriptive data, without an in-depth analysis of the type of BCTs used in the interventions, the use of behavior change theories to guide the interventions, the frequency with which each BCT was used, and potential associations between BCTs and intervention effectiveness. Therefore, this scoping review aims to analyze the use of BCTs in behavior change interventions delivered by CAs; specifically, it describes the health behaviors and disorders targeted by the intervention, describes the types of CAs used to deliver the behavior change interventions, identifies the theories or frameworks guiding the design of the behavior change interventions, identifies the most common type of BCTs used in CA-delivered interventions in health care, compares the BCTs employed in different types of CAs and for different health disorders, and compares the BCTs employed in the experimental and comparison interventions of studies evaluating CA-delivered behavior change interventions.

Methods

Overview

The scoping review was performed according to the Joanna Briggs Institute guidelines [18] and reported in alignment with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines (Multimedia Appendix 1) [19]. The protocol was registered in Open Science Framework Registries [20] in April 2021 and was published in a peer-reviewed journal in July 2021 [21].

Search Strategy

The search strategy was designed using a comprehensive list of words and phrases that define CAs (Multimedia Appendix 2). We searched PubMed, Embase (Ovid), and CENTRAL (Cochrane Central Register of Controlled Trials), from their inception, and the first 10 pages of Google and Google Scholar [22,23] on April 26, 2021.

Eligibility Criteria

This scoping review included primary, experimental studies in English evaluating the use of CAs to deliver health care interventions focusing on behavior change. Eligible study designs were randomized controlled trials (RCTs), quasi-RCTs, cluster-randomized trials, controlled before-and-after studies, uncontrolled before-and-after studies, interrupted time series, and pilot and feasibility studies. We excluded nonexperimental study designs, such as observational studies, qualitative studies, opinion pieces, editorials, conference abstracts, and secondary studies.

We included studies on text-based, voice-based, and embodied CAs, defined as conversational interfaces featuring a human-like

avatar able to mimic the verbal and nonverbal components of a face-to-face conversation [24]. The eligible studies reported any health care intervention focused on behavior change to improve or promote a healthy lifestyle, or to support the management of physical or mental health conditions. Lastly, behavior change was an essential aspect of the eligible studies, with or without reference to an associated behavior change theory, in line with previous research in this area [25]. The BCTs were coded according to the BCTTv1 [14]. The taxonomy consists of 93 BCTs grouped into 16 distinct categories, aimed at providing a cross-domain template to facilitate research and intervention replication.

Screening, Data Extraction, and Analysis

Screening

Screening for eligibility was performed in 2 stages. First, 2 researchers (NYWL and WWTG) worked independently to screen the titles and abstracts of all retrieved studies using Covidence [26]. Studies were excluded if their focus or study design did not align with our predefined eligibility criteria. Studies included in the first round of screening were uploaded to EndNote X9 (Clarivate), and the full-text papers were retrieved and screened for eligibility by 3 researchers working independently (AIJ, NYWL, and WWTG). Discrepancies in any screening stage were resolved through discussions between the reviewers, or by engaging a fourth reviewer (LM). The search and screening processes were documented in a study selection flowchart [27].

Data Extraction

The data were extracted using a Microsoft Excel (Microsoft Corporation) form developed by the research team, based on a data extraction form used in a previous scoping review [2], and a section on behavior change was added. The form was piloted in 3 studies and amended according to team members' feedback before being used for data extraction. Reviewers worked in pairs (AIJ worked with LM and NYWL worked with WWTG) to extract data from 10 papers (20%) and individually for the remaining 42 papers (80%). Data extracted by all reviewers were subsequently reexamined by 2 researchers (LM and AIJ). Reviewers met regularly during this process to ensure a common understanding of the data extraction process and the concordance of the extracted data. The data extracted by each pair of reviewers were compared, and any disagreements were resolved through consensus or consultation with a third reviewer, acting as an arbiter.

The data extraction form contained the following items: first author, year of publication, title of the article, study design, target disorder, description of the behavior change intervention, CA name, delivery channel, dialog technique, input and output modalities, end goal of the intervention, use of behavior change theories or frameworks, and BCTs mapped according to the BCTTv1 [14].

Data Analysis

Data were analyzed using descriptive statistics and frequent itemset mining (FIM) to explore possible BCT clustering [28]. Data were presented in a diagrammatic or tabular form accompanied by a narrative summary.

Frequent Itemset Mining

The FIM analysis was performed by implementing the Apriori algorithm using the *arules* package version 1.7-1 [29] in R version 4.1.2 (R Foundation for Statistical Computing) [30]. FIM aims to find patterns or associations in a group of items (itemset) by sorting the items that frequently appear together in the data set. The analysis starts by calculating support (how frequently an item appears in the data set) and confidence (number of times individual items "x" and "y" appear together in the data set) thresholds and discarding any itemset with support or confidence values below the predetermined minimum threshold.

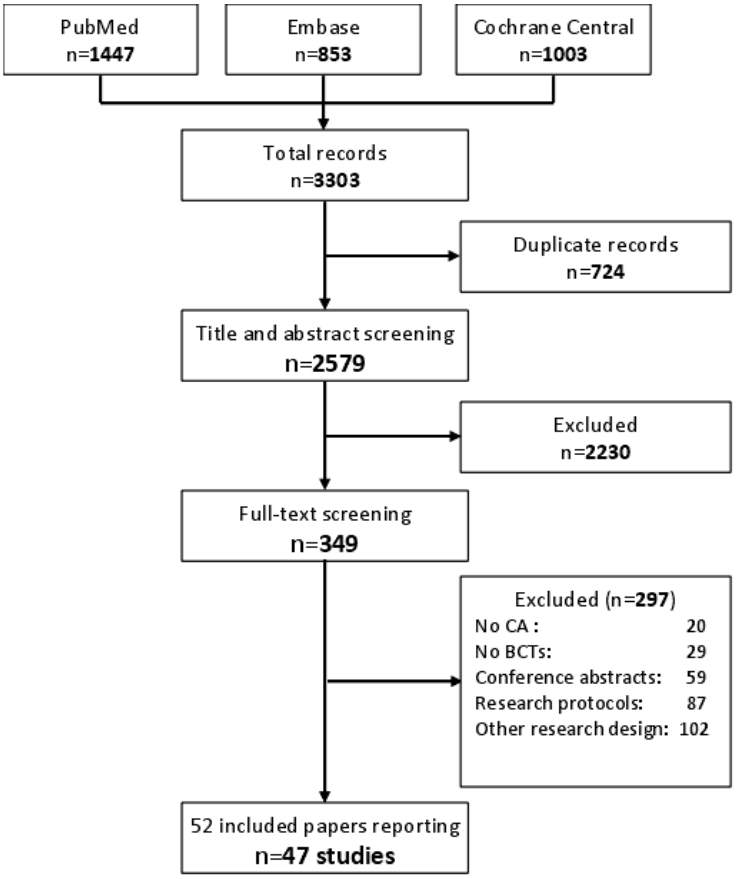
For this analysis, we assessed the 10 most frequently appearing patterns, for the overall data set and for each clinical domain. For the overall data set, the minimum threshold for algorithm support and confidence was set at 0.10 and 0.90, respectively, or itemset appearing in at least 10% of the data set (≥4 studies) and appearing together at least 90% of the time. For each clinical domain, the minimum thresholds were 0.20 for support and 0.90 for confidence to account for the fewer number of studies in each sub data set [31].

Results

Overview of Search Strategy

The search strategy retrieved 2579 papers after removing duplicates, of which 349 were eligible for full-text screening. Among these, 52 papers were finally included in this review. We reported 47 studies, as 4 studies were reported in 2 papers each and 1 study included a corrigendum. Figure 1 presents the study selection process.

Figure 1. Study selection flowchart. BCT: behavior change technique; CA: conversational agent.



Characteristics of the Included Studies

Multimedia Appendix 3 presents a summary of the studies included in this review [6,11,32-79]. Over half of the studies (26/47, 55%) were published from 2019 onward [11,32,34,37,40,42-46,48-55,58-61,65,66,71,72,76-78], including 6 published in the first quarter of 2021 [42,46,49,54,55,60]. All papers except 1 [32] were published in high-income countries, and 24/47 studies (51%) were published in the United States [6,32,34,36,39,43,45,47,48,51,52,54,56-58,61-64,67,69-75].

Most studies included a control group except 5/47 (11%) single-group pretest posttest trials [43,46,55,58,65,66], 3/47 (6%) feasibility studies [59-61], and 1/47 (2%) pilot study [48].

A total of 26/47 studies (55%) were RCTs [6,11,33,35-37,39-41,44,45,49,50,53,54,62-64,68-75,77,78]. In 36/47 studies (77%), the primary outcomes were associated with improvement of the target disorder [6,33,36,38-45,47-59,62-64,67,68,70-75,77-80], 5/47 studies (11%) reported technical-related primary outcomes (eg, technical performance, system crashes) [11,60,65,66,69,76], and 6/47 studies (13%) reported primarily user experience outcomes (eg, engagement with the CA, user satisfaction) [32,34,35,37,46,61]. Most interventions aimed to support treatment or monitoring (22/47, 47%) [6,33,35-44,46,48-50,53-55,59,60,80] or to promote healthy lifestyle change (18/47, 38%) [11,32,34,45,61-66,68-76,78,79]. Table 1 presents a summary of the included studies.

Table 1. Characteristics of included studies (N=47).

Study characteristics	Studies, n (%)
Year of publication	
Before 2019	21 (45)
2019 or after	26 (55)
Country	
United States	24 (51)
United Kingdom	6 (13)
Japan	3 (6)
Korea	3 (6)
Switzerland	3 (6)
Australia	2 (4)
France	1 (2)
Germany	1 (2)
India	1 (2)
Netherlands	1 (2)
Spain	1 (2)
Sweden	1 (2)
Study design	
Randomized controlled trial	26 (55)
Pilot study	9 (19)
Single-group pretest posttest trial	5 (11)
Feasibility study	5 (11)
Microrandomized controlled trials	1 (2)
Nonrandomized comparison study	1 (2)
Study outcomes	
Clinical	23 (49)
Clinical; user experience	12 (26)
User experience; clinical	6 (13)
Technical; clinical	3 (6)
Technical; clinical; user experience	2 (4)
Clinical; technical	1 (2)
Clinical focus of the interventions	
Lifestyle behavior change	17 (36)
Treatment and monitoring	16 (34)
Treatment and monitoring + education	4 (9)
Education	4 (9)
Education + lifestyle behavior change	3 (6)
Treatment and monitoring + lifestyle behavior change	2 (4)
Education + treatment and monitoring	1 (2)
Lifestyle behavior change + education	1 (2)
Clinical domains	
Mental health	19 (40)
Chronic disorders	14 (30)

Study characteristics	Studies, n (%)
Lifestyle modification	14 (30)

Clinical Domains

Mental Health Interventions

Most CAs focused on mental health (19/47, 40%) [6,32-47,79,80], either supporting mental well-being (5/19, 26%) for healthy individuals [46,47,79,80] or patients recovering from cancer [33]; enabling self-improvement interventions such as problem solving [34] or communications skills [35]; or assisting participants in the management of a mental health disorder (14/19, 74%) [6,36-46], including depression (with or without anxiety; 3/19, 16%) [6,36,37], emotional distress (2/19, 11%) [38,39], bipolar disorder [40], panic disorder [41], fear of heights [42], adult attention deficit disorder [43], substance use disorder [44], gambling [45], and social exclusion [46].

All except 2 interventions [44,47] included a control group, and 10/19 studies (53%) were RCTs [6,33-37,39,41,45,46]. A total of 6 studies included an active comparison with another digital intervention [34,38,39,46], a paper-based version of the CA intervention [40], or mood monitoring [33]. Besides, 6 studies provided information about the target disorder [6,35,37,41,43,48], and 10 experimental interventions (10/17, 59%) were reported as more effective than the comparisons [6,33-37,39,41,45,46].

Chronic Disorder Management Interventions

A total of 14/47 studies (30%) offered interventions focusing on a chronic disease other than mental illness [49-63]. Most studies (4/14, 29%) targeted a metabolic disorder including obesity (n=1) [63], prediabetes (n=1) [62], or type 2 diabetes (n=2) [51,56]. Three studies evaluated a pain management intervention for osteoarthritis (n=2) [57,58] or for general management of chronic pain (n=1) [54]. Other studies focused on asthma [61], atrial fibrillation [52,53], HIV [49], hypertension [50], insomnia [60], irritable bowel syndrome [55], and prostate cancer [59]. The interventions aimed to support treatment and monitoring tasks (8/14, 57%) or provide education (4/14, 29%).

Half of the included studies were feasibility or pilot studies, and 5/14 studies (36%) were RCTs [49,50,53,54,62]. Comparison interventions included a nurse-led instruction mirroring the CA intervention [50], physical activity monitoring using a pedometer [63], provision of information [57,58], treatment as usual [51-53], and waitlist controls [54,55]. Furthermore, 6/14 studies (43%) were single-group interventions without a comparison group [48,55,58-61]. Only 2 studies described the experimental interventions as more effective than the comparisons (2/8, 25%) [51,52,54].

Lifestyle Change Interventions

A total of 14/47 studies (30%) included interventions to support lifestyle modification [11,64-79], particularly increasing physical activity (10/14, 71%), either as the sole intervention (n=6) [64,69,74-77,79] or in combination with another approach such as diet improvement (n=2) [65-67], or diet improvement plus stress relief (n=1) [70]. Four studies (4/14, 29%) targeted an aspect of women's health including preconception care (n=3) [71-73,78] and breastfeeding support (n=1) [68]. One study offered a smoking cessation intervention [11]. In 12 studies, the interventions aimed to facilitate lifestyle change (12/14, 86%) [11,63-76,78], while 2 studies offered education [67,77].

Among this, 1/14 (7%) study was a single-group pretest-posttest trial [65,66], while most studies (11/14, 79%) were RCTs [11,63,64,68-75,77,78]. In 7/13 studies (54%) comparison interventions consisted of face-to-face versions of the intervention [74-76], abridged interventions that excluded the CA [11,64,65,70], or a similar version of the intervention with differing reward systems [77,79]. Other comparisons included information-only interventions (3/13, 23%), treatment as usual (1/13, 8%), or waitlists (2/13, 15%). Most experimental interventions were reported to be more effective than the comparisons (9/13, 69%).

Characteristics of CAs

Table 2 summarizes the characteristics of the included CAs.

A total of 39 CAs were included. Six CAs were reported in 2 or more manuscripts. Four CAs (Carmen [74-76], Tanya [52,53,68], Tess [37,62], and Todaki [41,43]) were reported in 2 papers each, and 2 CAs (Gabby [70-73] and MYLO [34,38,39]) were reported in 3 manuscripts. Three CAs were adapted for different target disorders. Embodied CA Tanya was used as an educational tool for patients with atrial fibrillation [52,53] and to offer breastfeeding support [68], CA Tess was used for mental health [37] and diabetes care [62], and Todaki was used to deliver CBT for panic disorder [41] and to manage adults with attention deficit disorder [43]. Finally, MYLO was used in student and older adult [38] populations by 2 distinct research groups.

The majority of CAs featured 1 or more anthropomorphic characteristics, such as the assignation of gender, name, or a human-like display. Most CAs (41/47, 87%) responded to a name, 27/47 CAs (57%) were presented as female agents, and 20/47 (43%) were embodied CAs. Most CAs used rule-based algorithms to design the flow of conversations, either by themselves (35/47, 75%) or complemented with AI (2/47, 4%). CAs were more often available through a smartphone app (14/47, 30%) or web page (13/47, 28%). In all but 3 CAs (44/47, 94%), the primary method for users' inputs was text; 7/47 of these CAs (15%) also accepted verbal or visual inputs, whereas 3/47 CAs (6%) received only verbal inputs. Almost 80% of all CAs (36/47, 77%) displayed a "coach-like" personality, characterized by an encouraging, motivating, and nurturing conversational style.

Table 2. Characteristics of CAs[a] (N=47).

CA characteristics	Values, n (%)
Type of CA	
Embodied CAs	20 (43)
No visual representation	12 (26)
Human-like cartoon avatar	10 (21)
Nonhuman cartoon avatar	5 (11)
Gender	
Female	27 (57)
No gender assigned (no avatar/no human avatar)	16 (34)
Male	2 (4)
Defined by the user	2 (4)
CA "level of intelligence"	
Rule-based CAs	34 (72)
Artificial intelligence CAs	9 (19)
Rule-based + artificial intelligence CAs	4 (9)
Dialog modality	
Predetermined text	28 (60)
Free text	8 (17)
Predetermined and free text	7 (15)
Not specified	4 (9)
Delivery channel	
Smartphone app	14 (30)
Web based	13 (28)
Desktop	7 (15)
Messaging apps	6 (13)
Two or more delivery channels	6 (13)
Tablet computer	1 (2)
Users' input modalities	
Text	37 (79)
Text + others (voice, images, video)	7 (15)
Voice (± video)	3 (6)
CA output modalities	
Text + others (voice, images, video)	29 (62)
Text	15 (32)
Voice (± images, video)	3 (6)
CA personality	
Coach like	36 (77)
Health care professional like	9 (19)
Not specified	2 (4)

[a]CA: conversational agent.

Type of CA and Clinical Domains

Embodied CAs were used to deliver almost two-thirds (9/14, 64%) of the interventions promoting lifestyle modification [64,65,68-76], 43% (6/14) of the chronic disease management interventions [49,51-53,59,60,63] and only 26% (5/19) of the mental health interventions.

By contrast, most mental health CAs did not include an avatar (8/19, 42%) [34,35,38-40,45,47,81], or they were represented by a nonhuman avatar (5/19, 26%) [6,33,41,43,44]. Human-like avatars were present in 1/19 (5%) mental health intervention [37], 6/14 (43%) chronic disease management interventions [54,55,57,58,61,62], and 3/14 (21%) lifestyle change interventions [66,67,77,78].

Behavior Change Theories and Techniques

Behavior Change Theories

A total of 12/47 (26%) studies incorporated a behavior change theory to guide the CA intervention design, including 4/14 (29%) studies targeting a chronic disorder [51,54,59,61], 7/14 (50%) studies [65,71-76,78,79] evaluating a lifestyle change intervention, and 1/19 study (5%) [37] on mental health. The Transtheoretical Model was the most used behavior change theory, either alone [37,71-73,78] or together with the Social Cognitive Theory [51,65,74-76]. In addition, 4/19 (21%) mental health studies and 2/14 (14%) studies targeting a chronic disorder based their interventions on theories derived from the behavior [34,38,39], communication [57,58], learning [59], or psychological domains [33] (Table 3).

The use of theories aimed to guide the design of the intervention or to monitor participants' stages of change as they progressed through the intervention, as exemplified by 3 studies [71-73,78] using the Transtheoretical Model and 1 study using the Health Action Process Approach [54]. It was not clear how the use of theories influenced the intervention design or the choice of BCTs. For example, 4 studies using the Transtheoretical Model included a wide variety of BCTs, ranging from 3 [78] to 10 [72,73]. Similarly, 4 studies [51,65,74-76] using the Transtheoretical Model and the Social Cognitive Theory incorporated between 6 [51] and 19 [75,76] BCTs.

Table 3. Behavior change theories informing the CA[a]-based interventions (N=47).

Theories guiding CA interventions	Studies, n (%)
No theory	29 (62)
Behavior change theories	11 (23)
Transtheoretical Model	4 (9)
Transtheoretical Model + Social Cognitive Theory	4 (9)
Theory of Planned Behavior + Self-Determination Theory + Technology Acceptance theories	1 (2)
Health Action Process Approach	1 (2)
Habit Formation Model	1 (2)
Behavior change theories + other theories	1 (2)
Unified Theory of Acceptance and Use of Technology + Cognitive Theory Multimedia Learning	1 (2)
Other theories	6 (13)
Perceptual Control Theory	3 (6)
Communication Accommodation Theory	2 (4)
Stress and Coping Theory + Broaden and Build Theory of Positive Emotion	1 (2)

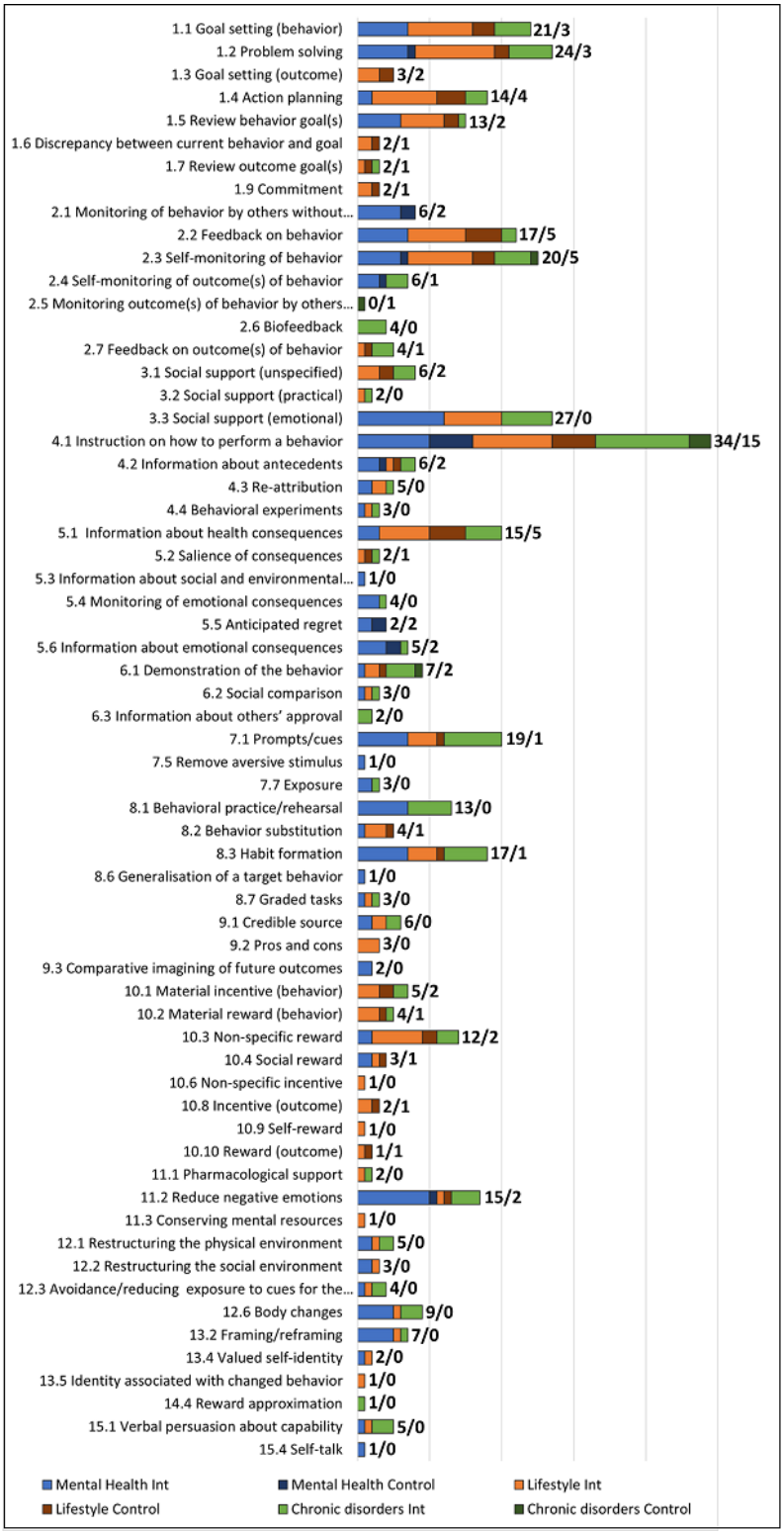
[a]CA: conversational agent.

Incorporated BCTs

The experimental interventions incorporated 63 BCTs from 15 categories, whereas the comparison interventions included 32 BCTs from 10 categories. However, only 24 BCTs were incorporated into experimental interventions in 5 or more studies, whereas 12 BCTs were reported in only 1 study each. The most incorporated BCT across interventions was 4.1 "Instruction on how to perform a behavior" (34/47, 72%), followed by 3.3 "Social support (emotional)" (27/47, 57%) and 1.2 "Problem solving" (24/47, 51%), whereas only 1 study included a BCT from category 14 (14.4 "Reward approximation") in the experimental intervention, and none included BCTs from category 16 "Covert learning." Figure 2 shows the frequency of presentation of all 63 BCTs in experimental and comparison interventions.

The average number of BCTs included in the experimental interventions was 9 (range 2-21 BCTs). By contrast, comparison interventions (n=38) included an average of 2 BCTs (range 0-17 BCTs).

Figure 2. Number of studies using each BCT in the experimental and comparison interventions. BCT: behavior change technique; Int: intervention.



Use of BCTs According to the Clinical Domain

The number of BCTs in experimental interventions was consistent across all clinical domains. Mental health interventions included an average of 8 BCTs (range 3-16 BCTs), chronic disorder management interventions included an average of 9 BCTs (range 2-18 BCTs), and lifestyle change interventions included an average of 10 BCTs (range 3-21 BCTs). The number of BCTs included in comparison interventions varied from an average of 2 BCTs in chronic disorder management (range 1-3 BCTs) and mental health interventions (range 1-2 BCTs) to a mean of 6 BCTs (range 1-17 BCTs) in lifestyle change interventions.

Mental health interventions incorporated 41 BCTs in experimental interventions. The most common BCTs were 3.3 "Social support (emotional)" (12/19, 63%), 11.2 "Reduce negative emotions" (11/19, 58%), 4.1 "Instruction on how to perform a behavior" (9/19, 47%), and BCTs 1.1 "Goal setting

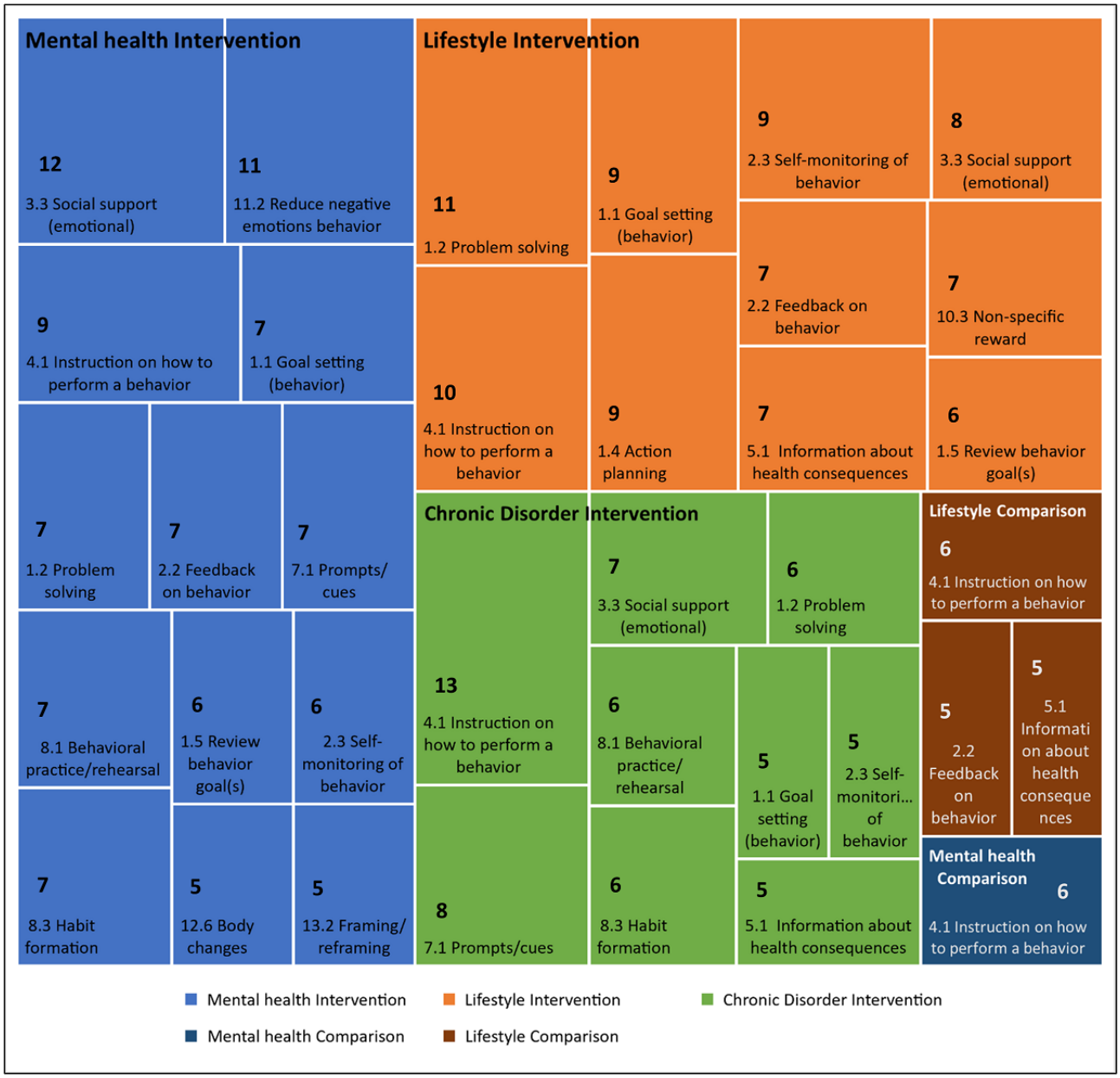
(behavior)," 1.2 "Problem solving," 2.2 "Feedback on behavior," 7.1 "Prompts/cues," 8.1 "Behavioral practice/rehearsal," and 8.3 "Habit formation" that were included in 7/19 (37%) studies each.

Lifestyle change interventions included 46 BCTs. The most common BCT was 1.2 "Problem solving" (11/14, 79%), followed by 4.1 "Instruction on how to perform a behavior" (10/14, 71%) and BCTs 1.1 "Goal setting (behavior)," 1.4 "Action planning," and 2.3 "Self-monitoring of behavior," included in 9/14 (64%) studies each.

Chronic disorder management interventions included a total of 41 BCTs. Almost all studies included BCT 4.1 "Instruction on how to perform a behavior" (13/14, 93%), followed by 7.1 "Prompts/cues" (8/14, 57%), 3.3 "Social support (emotional)" (7/14, 50%), and BCTs 1.2 "Problem solving," 8.1 "Behavioral practice/rehearsal," and 8.3 "Habit formation," all included in 6/14 studies (43%).

Figure 3 presents a summary of the most commonly used BCTs according to the clinical domain. Multimedia Appendix 4 presents a table summarizing the use of each BCT according to the clinical domain.

Figure 3. Commonly used BCTs according to the clinical domain. BCT: behavior change technique.



BCT Clustering According to the Clinical Domain Using FIM

The overall data set (n=47) generated 206 rules with an average support of 0.12, suggesting that the rules applied to at least 12% of the data set or about 6 studies. In general, 26% of the studies included BCTs 4.1 "Instruction on how to perform a behavior" and 8.1 "Behavioral practice/rehearsal," whereas 23% of the studies included BCTs 4.1 "Instruction on how to perform a behavior," 7.1 "Prompts/cues," and 8.3 "Habit formation."

The mental health domain (n=19) generated 45 rules with an average support of 0.22. About one-quarter of studies (26%) included 1 of 3 rules: the first itemset included BCTs 1.5

"Review behavior goal(s)," 2.2 "Feedback on behavior," and 3.3 "Social support"; followed by the itemset comprising BCTs 3.3 "Social support" and 12.6 "Body changes"; and the itemset containing BCTs 3.3 "Social support," 4.1 "Instruction on how to perform a behavior," and 11.2 "Reduce negative emotions." Conversely, the lifestyle change domain (n=14) generated 1322 rules with an average support of 0.24. About 64% of the studies included BCTs 1.2 "Problem solving" and 2.3 "Self-monitoring of behavior," whereas 57% of the studies also included BCT 1.1 "Goal-setting (behavior)." Finally, the chronic disorder management domain (n=14) generated 230 rules with an average support of 0.23. Most studies (93%) included BCT 4.1 "Instruction on how to perform a behavior," whereas 57% also included BCT 7.1 "Prompts/cues."

Multimedia Appendix 5 presents a table describing the top 10 itemsets for all included papers and each clinical domain.

Use of BCTs According to the CA Type

Interventions delivered by any type of CA included an average of 9 BCTs. However, the number of BCTs in experimental interventions varied by type of CA: embodied CAs included 2-19 BCTs, CAs represented by an avatar included 3-14 BCTs, and CAs with nonspecified or nonvisual representation incorporated 4-21 BCTs.

Embodied CAs included a total of 49 BCTs in the interventions. The most common BCTs were 3.3 "Social support (emotional) (14/20, 70%), and BCTs 1.2 "Problem solving," 2.3 "Self-monitoring of behavior," and 4.1 "Instruction on how to perform a behavior," which were found in 13/20 (65%) studies each. By contrast, CAs represented by an avatar included a total of 38 BCTs in the interventions. The most common BCTs were 4.1 "Instruction on how to perform a behavior" (13/15, 87%), and BCTs 3.3 "Social support (emotional)" and 7.1 "Prompts/cues" included in 10/15 (67%) studies each. Finally, CAs with nonspecified or nonvisual representation incorporated a total of 47 BCTs. Four BCTs (1.2 "Problem solving," 4.1 "Instruction on how to perform a behavior," 7.1 "Prompts/cues," and 8.3 "Habit formation") were included in 6/12 (50%) studies, and BCT 11.2 "Reduce negative emotions" was included in 5/12 (42%) studies. Multimedia Appendix 6 provides further information about the use of BCTs according to the type of CA.

Discussion

Principal Findings

This scoping review included 47 studies reporting behavior change interventions delivered by CAs, targeting chronic disorders, lifestyle change, and mental health. The interventions included a total of 63 BCTs, but only 24 were consistently found in 5 or more interventions. The BCTs represented aspects of health education (BCT 4.1), self-management (BCTs 1.1, 1.2, and 2.3), and social support (BCT 3.3). Several behavior change theories informed the intervention design in 12/47 (26%) studies of the included studies. However, studies informed by the same theory employed different sets of BCTs. Our findings align with previous systematic reviews reporting that similar BCTs were frequently incorporated into effective lifestyle change interventions [82], or into digitally delivered interventions [15].

We did not find a relationship between the use of theories, the type of theory used, and the number and type of BCTs included in the interventions. Furthermore, a small number of studies [11,61] guided the intervention design, using modified BCT taxonomies that addressed smoking cessation [11] and diet modification [61]. These data suggest that the choice of BCTs may be primarily determined by the target behavior rather than the use of a behavior change theory. The impact of using a behavior change theory is nevertheless unclear. A 2010 systematic review [83] reported that the use of a behavior change theory was associated with increased effectiveness of the interventions, although just over 20% of studies included a theory. Conversely, a systematic review by Van Rhoon et al [15] reported the use of theories in 16/21 (76%) studies but did not assess intervention effectiveness. In addition, a recent overview of systematic reviews [84] reported the use of theories in the intervention design of 19%-52% of the included studies, although there was no clear association with the intervention effectiveness.

The categorization of studies in 3 distinct clinical domains suggested different prioritizations in mental health, lifestyle change, and chronic disorders, although the delivery of health education, evidenced by the frequent occurrence of BCTs 4.1 "Instruction on how to perform a behavior," 8.1 "Behavioral practice/rehearsal," and 8.3 "Habit formation," was consistent across all clinical domains.

Mental health interventions frequently included BCTs 3.3 "Social support (emotional)" and 11.2 "Reduce negative emotions." Specifically, BCT 3.3 may be associated with the use of psychotherapeutic techniques such as cognitive behavioral therapy or motivational interviewing, while the inclusion of BCT 11.2 suggests the use of relaxation techniques and mindfulness to support stress management and emotional regulation. Therefore, behavior change in mental health settings appeared to be closely interlinked with the therapeutic strategies. Concurrently, the inclusion of other BCTs, such as instructions to perform a behavior (BCT 4.1), goal setting (BCT 1.1) and reviews (BCT 1.5), problem solving (BCT 1.2), and feedback (BCT 2.2), may be aligned with general principles of patient participation in decision making [85], as well as highlight the importance of health education [86,87], particularly relevant in self-initiated digital interventions.

Lifestyle change interventions frequently included problem-solving (BCT 1.2) techniques to help users better understand their barriers to behavior change, and goal setting (BCT 1.1) and self-monitoring (BCT 2.3) to work toward the target behavior. These BCTs were often included together and this may suggest a synergistic relationship. At the same time, the importance of ensuring adequate health literacy to improve population outcomes was emphasized by the frequent inclusion of BCT 4.1 "Instruction on how to perform a behavior."

Chronic disorder management interventions favored not only the inclusion of instructional BCTs, such as guidance to perform a target behavior (BCT 4.1) but also reminders (BCT 7.1 "Prompts/cues") to facilitate the acquisition of new routines (BCT 8.3 "Behavioral practice/rehearsal"). Self-management of chronic illnesses is essential to ensure improved patient

outcomes and adequate quality of life but requires that individuals engage in a steep learning curve as they adapt to living with a long-term condition and develop new habits.

In general, the relationship between the number and type of BCTs and the effectiveness of the interventions was inconsistent and appeared to be determined by the clinical domain. Effective lifestyle change interventions tended to include a higher number of BCTs, a finding that was not replicated in the other clinical domains. At the same time, lifestyle change interventions were comparatively more effective than those in other clinical domains, particularly chronic disorders. Effective interventions in the lifestyle change and mental health domains frequently included BCTs related to goal setting and planning, timely provision of feedback, health education, and rewards on completed tasks. Previous studies reported varied results. A 2017 systematic review of 48 studies [82] evaluating the management of overweight and obesity in adults found small pooled effect sizes for short- and long-term diet and physical activity interventions. Effective interventions included a larger number of BCTs, particularly BCTs encouraging goal setting and self-monitoring of behavior. Similarly, a systematic review on the BCTs and technical features of digital interventions for the prevention of type 2 diabetes [15] found that effective interventions included a larger number of BCTs or BCTs related to social support, goal setting, and feedback.

There was an unexpected relationship between the CA types and the clinical domain, manifested by a predominance of embodied CAs in lifestyle change interventions, and the use of nonhuman or nonavatar CAs in mental health interventions. The reasons for these findings are unclear and beyond the scope of this review; however, further research may help clarify the role of avatars, or virtual humans, if any, in delivering behavior change interventions. Other reviews have reported the use of embodied CAs to support mental health interventions, particularly autism [20,24], but methodological differences limit the comparisons with our findings. Provoost et al's scoping review [4] used a broader definition of embodied CA, while a systematic review by Laranjo et al [87] included only AI-based CAs.

Strengths and Limitations

This scoping review has several strengths. First, we used a comprehensive literature search of peer-reviewed and gray literature that prioritized the sensitivity of the search terms to capture a broad range of publications reporting the use of CAs in health care. However, relevant studies may have been omitted. Second, we included studies reporting on a wide variety of physical and mental health conditions, and categorized the studies into 3 distinct clinical domains, revealing differences in the type of BCTs selected in each domain.

There are also some limitations. First, many studies did not provide exact BCT codes when describing the interventions, therefore categorization of BCTs was inferred from the paper's description by the research team, based on thorough analysis, rigorous team discussion, and reviews to establish consensus. Second, given the descriptive nature of scoping reviews, we were unable to explore in more depth the relationship between the choice of BCTs and the effectiveness of the intervention, or the type of CA used to deliver the intervention.

Future Research and Practice Recommendations

This review has highlighted several areas that warrant further research. First, reporting guidelines to ensure accurate reporting of the BCTs included in behavior change interventions according to standardized taxonomies, such as the BCTTv1 [14], should be implemented. Such guidelines would facilitate reproducibility of research, assessment of active intervention components, and evidence synthesis. Second, further research is needed to increase our understanding of the impact of behavior change theories in the design of interventions, the choice of BCTs, and the effectiveness of the intervention. Third, the impact of CAs to deliver behavior change interventions should be further explored, particularly the influence of a conversational interface on engagement, adherence, and effectiveness of the intervention when compared with less interactive digital technologies. Furthermore, comparisons between rule-based CAs and those incorporating machine learning or natural language processing should be further investigated. Fourth, the possible role of the type of CA in delivering behavior change interventions, as suggested in our findings, should be further explored. Fifth, the relationship between the ideal combination of BCTs required to design effective interventions may be evaluated using data mining techniques such as FIM or multiple correspondent analysis. Lastly, the relationship between behavior change interventions and mental health requires further evaluation.

The use of CAs to deliver behavior change interventions appears promising, particularly to support lifestyle change, although better reporting of BCTs included in the interventions is warranted to facilitate analysis of active components, design more effective interventions, and ensure reproducibility of research. The role of CA types in delivering behavior change interventions should be further explored.

Acknowledgments

This research is supported by the Singapore Ministry of Education under Singapore Ministry of Education Academic Research Fund Tier 1 (RG36/20). The research was conducted as part of the Future Health Technologies program, which was established collaboratively between ETH Zurich and the National Research Foundation, Singapore. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors' Contributions

LTC conceptualized the study and provided supervision at all steps of research. LTC and LM designed the study. LM, AIJ, WWTG, and NYWL extracted data and conducted the analysis. LM and AIJ wrote the manuscript. MHRH, TK, RA, and SM provided critical review of the manuscript. All authors approved the final version of the manuscript and take accountability for all aspects of the work.

Conflicts of Interest

TK is affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St.Gallen, which is funded in part by CSS, a Swiss health insurer. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies was involved in this research. The other authors declare that they have no competing interests.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - jmir_v24i10e39243_app1.docx]

Multimedia Appendix 2
PubMed search strategy.
[DOCX File , 17 KB - jmir_v24i10e39243_app2.docx]

Multimedia Appendix 3
Characteristics of included studies.
[DOCX File , 42 KB - jmir_v24i10e39243_app3.docx]

Multimedia Appendix 4
Use of BCTs according to the clinical domain. BCT: behavior change technique.
[DOCX File , 25 KB - jmir_v24i10e39243_app4.docx]

Multimedia Appendix 5
Frequent Itemset Mining (FIM).
[DOCX File , 20 KB - jmir_v24i10e39243_app5.docx]

Multimedia Appendix 6
Use of BCTs according to the CA type. BCT: behavior change technique; CA: conversational agent.
[DOCX File , 426 KB - jmir_v24i10e39243_app6.docx]

References

1. Conversational agents or chatbots. Oxford Learner's Dictionaries. URL: https://www.oxfordlearnersdictionaries.com/definition/english/chatbot [accessed 2021-03-23]
2. 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]
3. Diederich S, Brendel A, Kolbe L. Towards a Taxonomy of Platforms for Conversational Agent Design. In: Proceedings of Internationale Tagung Wirtschaftsinformatik. 2019 Presented at: Internationale Tagung Wirtschaftsinformatik; February 24-27, 2019; Siegen, Germany URL: https://www.researchgate.net/publication/329739579_Towards_a_Taxonomy_of_Platforms_for_Conversational_Agent_Design/
4. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
5. Amith M, Zhu A, Cunningham R, Lin R, Savas L, Shay L, et al. Early Usability Assessment of a Conversational Agent for HPV Vaccination. Stud Health Technol Inform 2019;257:17-23 [FREE Full text] [Medline: 30741166]
6. 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]
7. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]

8. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
9. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
10. Eisele A, Schagg D, Krämer LV, Bengel J, Göhner W. Behaviour change techniques applied in interventions to enhance physical activity adherence in patients with chronic musculoskeletal conditions: A systematic review and meta-analysis. Patient Educ Couns 2019 Jan;102(1):25-36. [doi: 10.1016/j.pec.2018.09.018] [Medline: 30279029]
11. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
12. 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]
13. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
14. 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]
15. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020 Dec;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
16. Pereira J, Díaz. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
17. 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]
18. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
20. The choice of behavioural change techniques in conversational agents in healthcare: a scoping review protocol. OSF Registry. 2021 Apr 28. URL: https://osf.io/487jd [accessed 2022-09-21]
21. Martinengo L, Lo NYW, Goh WIWT, Tudor Car L. Choice of Behavioral Change Techniques in Health Care Conversational Agents: Protocol for a Scoping Review. JMIR Res Protoc 2021 Jul 21;10(7):e30166 [FREE Full text] [doi: 10.2196/30166] [Medline: 34287221]
22. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]
23. Haddaway NR, Collins AM, Coughlin D, Kirk S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLoS One 2015;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
24. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
25. Covidence. URL: https://www.covidence.org/about-us/ [accessed 2021-03-28]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Luna JM, Fournier‐Viger P, Ventura S. Frequent itemset mining: A 25 years review. WIREs Data Mining Knowl Discov 2019 Jul 16;9(6):1329. [doi: 10.1002/widm.1329]
28. Hahsler M, Grün B, Hornik K. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. J. Stat. Soft 2005;14(15):1-25. [doi: 10.18637/jss.v014.i15]
29. R Foundation for Statistical Computing. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

30. Luna JM, Ondra M, Fardoun HM, Ventura S. Optimization of quality measures in association rule mining: an empirical study. IJCIS 2018;12(1):59. [doi: 10.2991/ijcis.2018.25905182]
31. World Bank Country and Lending Groups. World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2021-11-01]
32. 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]
33. 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]
34. Ali R, Hoque E, Duberstein P, Schubert L, Razavi SZ, Kane B, et al. Aging and Engaging: A Pilot Randomized Controlled Trial of an Online Conversational Skills Coach for Older Adults. Am J Geriatr Psychiatry 2021 Aug;29(8):804-815. [doi: 10.1016/j.jagp.2020.11.004] [Medline: 33308893]
35. Burton C, Szentagotai TA, McKinstry B, Matheson C, Matu S, Moldovan R, et al. 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]
36. 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]
37. 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 Mar 12:e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
38. 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]
39. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord 2015 Mar 15;174:23-30. [doi: 10.1016/j.jad.2014.10.053] [Medline: 25479050]
40. Oh J, Jang S, Kim H, Kim J. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. Int J Med Inform 2020 Aug;140:104171. [doi: 10.1016/j.ijmedinf.2020.104171] [Medline: 32446158]
41. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Dec;5(8):625-632. [doi: 10.1016/S2215-0366(18)30226-8] [Medline: 30007519]
42. 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]
43. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. J Med Internet Res 2021 Mar 23;23(3):e24850. [doi: 10.2196/24850]
44. So R, Furukawa TA, Matsushita S, Baba T, Matsuzaki T, Furuno S, et al. Unguided Chatbot-Delivered Cognitive Behavioural Intervention for Problem Gamblers Through Messaging App: A Randomised Controlled Trial. J Gambl Stud 2020 Dec;36(4):1391-1407. [doi: 10.1007/s10899-020-09935-4] [Medline: 32162075]
45. 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]
46. Bendig E, Erb B, Meißner D, Bauereiß N, Baumeister H. Feasibility of a Software agent providing a brief Intervention for Self-help to Uplift psychological wellbeing ("SISU"). A single-group pretest-posttest trial investigating the potential of SISU to act as therapeutic agent. Internet Interv 2021 Apr;24:100377 [FREE Full text] [doi: 10.1016/j.invent.2021.100377] [Medline: 33816127]
47. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ Couns 2013 Aug;92(2):160-166 [FREE Full text] [doi: 10.1016/j.pec.2013.05.007] [Medline: 23809167]
48. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, Feasibility, and Preliminary Efficacy of a Theory-Based Relational Embodied Conversational Agent Mobile Phone Intervention to Promote HIV Medication Adherence in Young HIV-Positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
49. Echeazarra L, Pereira J, Saracho R. TensioBot: a Chatbot Assistant for Self-Managed in-House Blood Pressure Checking. J Med Syst 2021 Mar 15;45(4):54. [doi: 10.1007/s10916-021-01730-x] [Medline: 33723721]
50. 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]

51. Guhl E, Althouse AD, Pusateri AM, Kimani E, Paasche-Orlow MK, Bickmore TW, et al. The Atrial Fibrillation Health Literacy Information Technology Trial: Pilot Trial of a Mobile Health App for Atrial Fibrillation. JMIR Cardio 2020 Sep 04;4(1):e17162 [FREE Full text] [doi: 10.2196/17162] [Medline: 32886070]
52. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The Atrial Fibrillation Health Literacy Information Technology System: Pilot Assessment. JMIR Cardio 2017;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
53. 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]
54. Hunt M, Miguez S, Dukas B, Onwude O, White S. Efficacy of Zemedy, a Mobile Digital Therapeutic for the Self-management of Irritable Bowel Syndrome: Crossover Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 May 20;9(5):e26152 [FREE Full text] [doi: 10.2196/26152] [Medline: 33872182]
55. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating Glycemic Control in Patients of South Asian Origin With Type 2 Diabetes Using a Digital Therapeutic Platform: Analysis of Real-World Data. J Med Internet Res 2021 Mar 25;23(3):e17908 [FREE Full text] [doi: 10.2196/17908] [Medline: 33764306]
56. McDonald DD, Walsh S, Vergara C, Gifford T. Effect of a virtual pain coach on pain management discussions: a pilot study. Pain Manag Nurs 2013 Dec;14(4):200-209. [doi: 10.1016/j.pmn.2011.03.004] [Medline: 24315243]
57. McDonald DD, Walsh S, Vergara C, Gifford T, Weiner DK. The effect of a Spanish virtual pain coach for older adults: a pilot study. Pain Med 2012 Nov;13(11):1397-1406. [doi: 10.1111/j.1526-4637.2012.01491.x] [Medline: 22994410]
58. Owens OL, Felder T, Tavakoli AS, Revels AA, Friedman DB, Hughes-Halbert C, et al. Evaluation of a Computer-Based Decision Aid for Promoting Informed Prostate Cancer Screening Decisions Among African American Men: iDecide. Am J Health Promot 2019 Feb;33(2):267-278. [doi: 10.1177/0890117118786866] [Medline: 29996666]
59. Philip P, Dupuy L, Morin CM, de Sevin E, Bioulac S, Taillard J, et al. Smartphone-Based Virtual Agents to Help Individuals With Sleep Concerns During COVID-19 Confinement: Feasibility Study. J Med Internet Res 2020 Dec 18;22(12):e24268 [FREE Full text] [doi: 10.2196/24268] [Medline: 33264099]
60. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational Agents as Mediating Social Actors in Chronic Disease Management Involving Health Care Professionals, Patients, and Family Members: Multisite Single-Arm Feasibility Study. J Med Internet Res 2021 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
61. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
62. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012 Jan 26;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
63. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
64. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
65. Davis CR, Murphy KJ, Curtis RG, Maher CA. A Process Evaluation Examining the Performance, Adherence, and Acceptability of a Physical Activity and Diet Artificial Intelligence Virtual Health Assistant. Int J Environ Res Public Health 2020 Dec 07;17(23):1-14 [FREE Full text] [doi: 10.3390/ijerph17239137] [Medline: 33297456]
66. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
67. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
68. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
69. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729. [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
70. Jack B, Bickmore T, Hempstead M, Yinusa-Nyahkoon L, Sadikova E, Mitchell S, et al. Reducing Preconception Risks Among African American Women with Conversational Agent Technology. J Am Board Fam Med 2015;28(4):441-451 [FREE Full text] [doi: 10.3122/jabfm.2015.04.140327] [Medline: 26152434]

71. Jack BW, Bickmore T, Yinusa-Nyahkoon L, Reichert M, Julce C, Sidduri N, et al. Improving the health of young African American women in the preconception period using health information technology: a randomised controlled trial. Lancet Digit Health 2020 Sep;2(9):e475-e485 [FREE Full text] [doi: 10.1016/S2589-7500(20)30189-8] [Medline: 33328115]
72. 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]
73. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464 [FREE Full text] [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
74. King AC, Campero I, Sheats JL, Castro Sweet CM, Garcia D, Chazaro A, et al. Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. Contemp Clin Trials 2017 Oct;61:115-125 [FREE Full text] [doi: 10.1016/j.cct.2017.07.020] [Medline: 28739541]
75. King AC, Campero MI, Sheats JL, Castro Sweet CM, Hauser ME, Garcia D, et al. Effects of Counseling by Peer Human Advisors vs Computers to Increase Walking in Underserved Populations: The COMPASS Randomized Clinical Trial. JAMA Intern Med 2020 Nov 01;180(11):1481-1490 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4143] [Medline: 32986075]
76. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
77. 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. Reproductive BioMedicine Online 2020 Dec;41(6):1133-1143. [doi: 10.1016/j.rbmo.2020.09.006] [Medline: 33039321]
78. 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]
79. 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]
80. 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]
81. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
82. 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]
83. Dalgetty R, Miller CB, Dombrowski SU. Examining the theory-effectiveness hypothesis: A systematic review of systematic reviews. Br J Health Psychol 2019 May;24(2):334-356. [doi: 10.1111/bjhp.12356] [Medline: 30793445]
84. Slade M. Implementing shared decision making in routine mental health care. World Psychiatry 2017 Jun;16(2):146-153 [FREE Full text] [doi: 10.1002/wps.20412] [Medline: 28498575]
85. Furnham A, Swami V. Mental Health Literacy: A Review of What It Is and Why It Matters. International Perspectives in Psychology 2018 Oct;7(4):240-257. [doi: 10.1037/ipp0000094]
86. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. Am Psychol 2012 Apr;67(3):231-243. [doi: 10.1037/a0025957] [Medline: 22040221]
87. 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]

Abbreviations

AI: artificial intelligence
BCT: behavior change technique
BCTTv1: Behavior Change Technique Taxonomy version 1
CA: conversational agent
CENTRAL: Cochrane Central Register of Controlled Trials
FIM: frequent itemset mining
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

RCT: randomized controlled trial

Edited by A Mavragani; submitted 04.05.22; peer-reviewed by L Van Rhoon, M Jalan; comments to author 01.07.22; revised version received 05.08.22; accepted 23.08.22; published 03.10.22.

Please cite as:
Martinengo L, Jabir AI, Goh WWT, Lo NYW, Ho MHR, Kowatsch T, Atun R, Michie S, Tudor Car L
Conversational Agents in Health Care: Scoping Review of Their Behavior Change Techniques and Underpinning Theory
J Med Internet Res 2022;24(10):e39243
URL: https://www.jmir.org/2022/10/e39243
doi:10.2196/39243
PMID:36190749

Original Paper

Secure Collaborative Platform for Health Care Research in an Open Environment: Perspective on Accountability in Access Control

Giluk Kang[1], BE; Young-Gab Kim[1], BS, MS, PhD

Department of Computer and Information Security, and Convergence Engineering for Intelligent Drone, Sejong University, Seoul, Republic of Korea

Corresponding Author:
Young-Gab Kim, BS, MS, PhD
Department of Computer and Information Security, and Convergence Engineering for Intelligent Drone
Sejong University
209, Neungdong-ro, Gwangjin-gu
Seoul, 05006
Republic of Korea
Phone: 82 0269352424
Email: alwaysgabi@sejong.ac.kr

Abstract

Background: With the recent use of IT in health care, a variety of eHealth data are increasingly being collected and stored by national health agencies. As these eHealth data can advance the modern health care system and make it smarter, many researchers want to use these data in their studies. However, using eHealth data brings about privacy and security concerns. The analytical environment that supports health care research must also consider many requirements. For these reasons, countries generally provide research platforms for health care, but some data providers (eg, patients) are still concerned about the security and privacy of their eHealth data. Thus, a more secure platform for health care research that guarantees the utility of eHealth data while focusing on its security and privacy is needed.

Objective: This study aims to implement a research platform for health care called the health care big data platform (HBDP), which is more secure than previous health care research platforms. The HBDP uses attribute-based encryption to achieve fine-grained access control and encryption of stored eHealth data in an open environment. Moreover, in the HBDP, platform administrators can perform the appropriate follow-up (eg, block illegal users) and monitoring through a private blockchain. In other words, the HBDP supports accountability in access control.

Methods: We first identified potential security threats in the health care domain. We then defined the security requirements to minimize the identified threats. In particular, the requirements were defined based on the security solutions used in existing health care research platforms. We then proposed the HBDP, which meets defined security requirements (ie, access control, encryption of stored eHealth data, and accountability). Finally, we implemented the HBDP to prove its feasibility.

Results: This study carried out case studies for illegal user detection via the implemented HBDP based on specific scenarios related to the threats. As a result, the platform detected illegal users appropriately via the security agent. Furthermore, in the empirical evaluation of massive data encryption (eg, 100,000 rows with 3 sensitive columns within 46 columns) for column-level encryption, full encryption after column-level encryption, and full decryption including column-level decryption, our approach achieved approximately 3 minutes, 1 minute, and 9 minutes, respectively. In the blockchain, average latencies and throughputs in 1Org with 2Peers reached approximately 18 seconds and 49 transactions per second (TPS) in read mode and approximately 4 seconds and 120 TPS in write mode in 300 TPS.

Conclusions: The HBDP enables fine-grained access control and secure storage of eHealth data via attribute-based encryption cryptography. It also provides nonrepudiation and accountability through the blockchain. Therefore, we consider that our proposal provides a sufficiently secure environment for the use of eHealth data in health care research.

(J Med Internet Res 2022;24(10):e37978) doi:10.2196/37978

KEYWORDS

blockchain; attribute-based encryption; eHealth data; security; privacy; cloud computing; research platform for health care; accountability; Internet of Things; interoperability; mobile phone

Introduction

Background

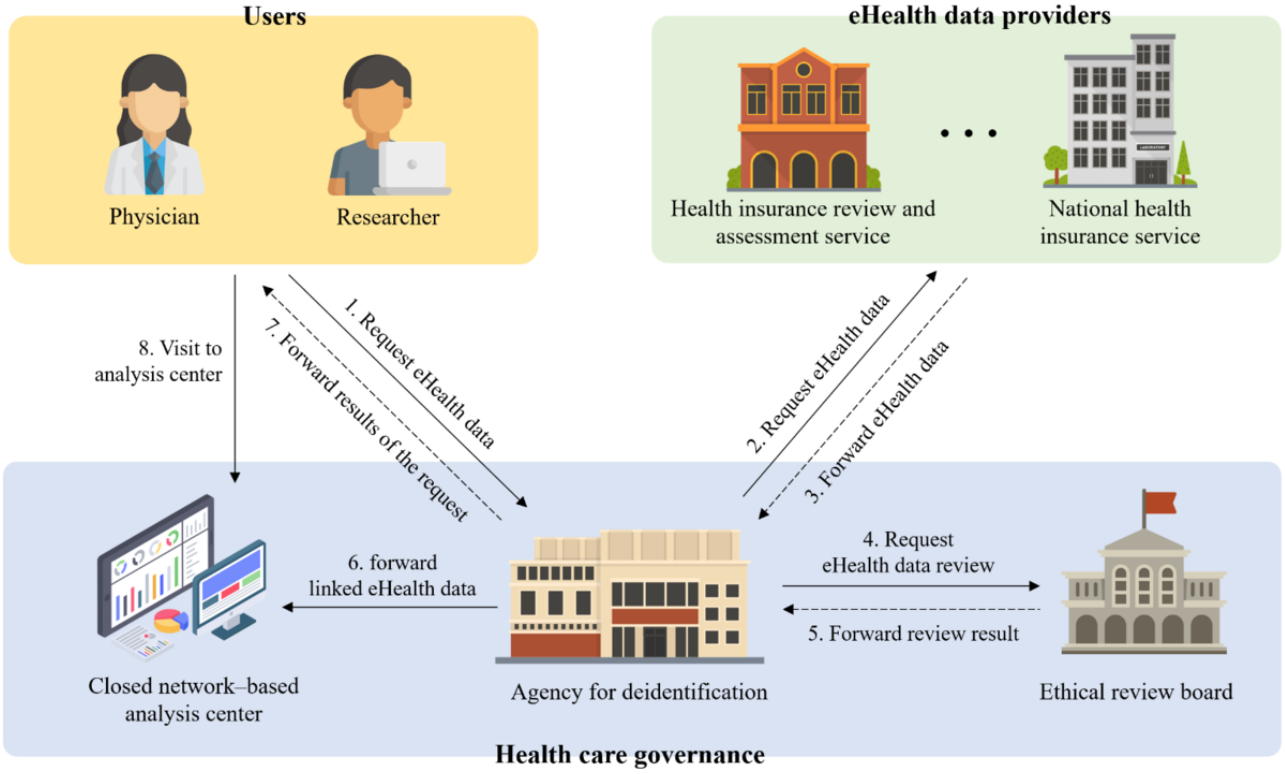
The development of modern technologies such as the Internet of Things (IoT), cloud computing, big data, and blockchain affects many aspects of human life. Primarily, these technologies have introduced changes in health care. The quality of health care services and operations has also improved because of the digitization of the health care system. Furthermore, with the advancement in sensors, the eHealth data generated by IoT devices for health care are increasingly being collected by health facilities and national health agencies. These eHealth data generally include electronic medical records (EMRs) and personal health records (PHRs), which contain a considerable amount of personal information such as any disease a patient may have and the patient's medical record number. Thus, some eHealth data subjects have expressed security and privacy concerns related to the use of eHealth data. For this reason, the use of eHealth data is currently governed by many legal regulations, including the Health Insurance Portability and Accountability Act [1], General Data Protection Regulation (GDPR) [2], and California Consumer Privacy Act [3]. However, the security of eHealth data has frequently been breached, and the number of cyberattacks launched to hijack eHealth data intended for health care services is on the rise [4].

Nevertheless, using eHealth data for health care research has many advantages (eg, improving treatment and prescriptions for patients, increasing the efficiency of health care systems, and expanding knowledge of diseases), so many researchers hope to use them for their studies [5]. However, the interoperability, utility, and data linkage of eHealth data as well as privacy laws (eg, Health Insurance Portability and Accountability Act, GDPR, and California Consumer Privacy Act) and analytics tools must be considered when a research platform for health care is built. Furthermore, security and privacy measures (eg, anonymization and access control) for an open research environment for eHealth data are needed, and many privacy laws must be complied with. Owing to these complex requirements, most research platforms for health care development are being led by national governments. For example, as depicted in Figure 1, the Ministry of Health and Welfare of South Korea [6] operates a closed network–based analysis center that supports a research environment for analyzing eHealth data. However, researchers must visit the analysis center as they are not able to connect to it remotely or on the web. Not only is this analysis center inconvenient to visit, but it also presents a challenge to efficiently analyzing eHealth data as programming errors can only be corrected via books because of the closed nature of the network. Moreover, the eHealth data requested by the researchers are immediately deleted after use, which reduces the utility of the data.

The National Health Service (NHS) in England also offers eHealth data to researchers and clinicians through a Data Access Request Service (DARS) [7]. The NHS DARS provides various analytical tools such as Databricks, R Studio, and Hue in the data access environment, and it does not require the researcher to visit the research analysis center, unlike the center in South Korea. The NHS DARS also provides many security solutions (eg, 2-factor authentication, data-sharing audits, and anonymization) to ensure the security and privacy of eHealth data. Furthermore, the Swiss Personalized Health Network offers a secure infrastructure for the exchange and use of eHealth data for research [8]. In the Swiss Personalized Health Network, eHealth data can be accessed only from reliable hospitals and universities or the virtual private network, which are environments. Researchers must take the web-based ethics training and are required to complete 2-factor authentication. However, data subjects (ie, patients) are still concerned about unauthorized data reuse and sharing, and they hope to be involved in eHealth data access decisions [9]. In addition, even if eHealth data are deidentified and anonymized, reidentification is still possible via other big data [10,11]. In other words, studies on health care research platforms are needed to provide a more secure analytical environment in light of the apprehension of data subjects regarding the security and privacy of their eHealth data.

Therefore, we propose a secure research platform for health care, referred to as the health care big data platform (HBDP). In this study, we considered only a secure and open research environment, although a research platform for health care has many requirements. The HBDP uses a private blockchain to provide a decentralized persistent log database (DB) in which user activities on the platform are recorded with a time stamp by a smart contract. This helps the platform administrator conduct the appropriate follow-up and monitoring of security threats. Furthermore, the HBDP uses attribute-based encryption (ABE) to ensure the security and privacy of eHealth data and prevent eHealth data leakage by insiders. To the best of our knowledge, this is the first study on a secure research platform that is focused on accountability to secure the use of eHealth data in an open environment based on blockchain and ABE. The main contributions of this study are summarized in Textbox 1.

Figure 1. Access procedure for analyzing eHealth data in South Korea.



Textbox 1. Main contributions of this study.

Main contributions

- We propose the health care big data platform (HBDP), which supports the accountability, access control, and encryption of stored eHealth data using attribute-based encryption and a private blockchain in an open environment. In particular, we focused on accountability in access control. We also analyzed previous research platforms for health care from a security perspective.
- For accountability in access control, a smart contract is designed to record in real time the success or failure of user activities (eg, log-in and use of eHealth data) on the HBDP. In particular, the contract enables user monitoring and illegal user detection in the HBDP anytime.
- To prove and demonstrate the feasibility of the HBDP, we implemented a framework for the HBDP using Hyperledger Fabric (The Linux Foundation) [12], OpenABE library (Zeutro) [13], and OpenStack (Open Infrastructure Foundation) [14], and we evaluated its security by using case studies on the detection of illegal users.

Prior Work

Overview

To analyze prior work, we first collected and analyzed well-known health care standards for the interoperability of eHealth data. After analyzing the standards, we searched existing health care studies related to the blockchain using the terms “blockchain” AND “access OR data sharing OR access control” AND “healthcare” for literature review in IEEE Xplore, Wiley Online Library, ScienceDirect, and MDPI. The results identified 501 papers in IEEE Xplore, 943 articles in the Wiley Online Library, 2599 articles in ScienceDirect, and 24,219 articles in MDPI. To select suitable studies, we added some filters (ie, published from 2018 to 2022 and cited by ≥5 journals) based on these results. We also reviewed the abstracts and titles of the papers. On the basis of these works, we finally selected 9 papers (ie, IEEE Xplore: n=4, 44%; Wiley Online Library: n=2, 22%; ScienceDirect: n=2, 22%; and MDPI: n=1, 11%).

Furthermore, we searched health care research platforms from 2015 to 2022 using the terms “healthcare research platform” and “clinical research platform” in Google Scholar. The results showed approximately 849,000 and 1,480,000 papers for each keyword, respectively. To identify suitable studies, we also reviewed the abstracts and titles. In particular, we examined the security solutions in each study and finally selected 6 papers. This section analyzes the identified studies via these processes in detail.

Standards for Interoperability of eHealth Data

For a long time, eHealth data have been limited to being shared and accessed between health care providers owing to interoperability issues such as differences in representation (eg, vocabularies and terminologies), equipment, and data formats. These issues currently make it difficult for health care providers to ensure continuity of care for patients or analyze eHealth data in health care. Therefore, many health care organizations are publishing interoperability standards for eHealth data in health care. Clinical Document Architecture (CDA) [15] is an XML-based markup standard for clinical document exchange

designed by Health Level 7 (HL7). CDA prescribes the structure and semantics of clinical documents for interaction between health care systems. The central aspect of CDA is easily exchanging clinical documents and making them readable. However, CDA-based documentation has the disadvantage of making it complex and difficult. For this reason, CDA has been extended to Consolidated CDA with improved complexity and interoperability. Fast Healthcare Interoperability Resources (FHIR) [16] is a standard to ensure the interoperability of health care systems or services also developed by HL7. The FHIR improved the limitations of the previously developed HL7 versions 2 and 3 (eg, implementation complexity and structured data model) to make the exchange of medical information easier. Furthermore, it was developed based on the representational state transfer architecture, so it is easy to implement health care services for mobile phones, wearable devices, and tablet devices beyond computers. Thus, the FHIR is currently one of the most popular standards for the interoperability of eHealth data in health care. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [17] is an open community standard for the eHealth data model managed by Observational Health Data Sciences and Informatics. The OMOP CDM solves the interoperability issues of eHealth by structuring the data model and the content of observational data. The OMOP CDM structures eHealth data to provide a common data model and converts them into a common representation through the OMOP to provide the common physical and logical interoperability model. When a health care DB is designed via the OMOP CDM, it can use standardized analysis tools and help analyze eHealth data systematically. It also increases the efficiency of joint research. Digital Imaging and Communications in Medicine (DICOM) [18] is a data format standard for the interoperability of medical imaging such as magnetic resonance imaging, computed tomography, and x-rays. DICOM has defined the format of medical imaging so that medical images captured by various imaging devices can be transmitted and exchanged. DICOM is generally stored, processed, and transmitted via the picture archiving and communication system and is the best known today in health care. Cross-Enterprise Document Sharing (XDS) [19] is an integrated profile for eHealth data developed by Integrating the Healthcare Enterprise in 2004. In particular, XDS can share various standard-based clinical documents such as the HL7 CDA, general strings, and binary data. In other words, XDS represents a comprehensive and universal technology. In addition to the aforementioned standards, various standards are being established for the interoperability of eHealth data by many health care organizations. We consider that these standards do not provide perfect interoperability of eHealth data but can still be addressed in the near future. Thus, the interoperability of eHealth data is the main requirement in the research platform for health care, but it is not the main focus of this study.

Secure eHealth Data Sharing via Blockchain

The blockchain has many advantages (eg, data integrity, decentralization, and programmable smart contract), so many research areas have been trying to use it. In particular, the blockchain has been widely used to address the integrity, scalability, and sharing of eHealth data. However, in addition, eHealth data require security mechanisms such as access control, cryptography, and authentication owing to privacy and security issues. For this reason, many studies generally use these security mechanisms with the blockchain. Table 1 shows the strengths and weaknesses of these studies and the HBDP. Yang et al [20] proposed an architecture that can use blockchain in the existing health care system. The architecture has recorded all accesses, such as select, insert, and delete, using two smart contracts (ie, summary contract and record relationship contract) to ensure the integrity of data records. The architecture also performs access control via an access control list. Madine et al [21] proposed a blockchain-based, patient-centric PHR management system. The system uses trusted oracles that perform proxy re-encryption to share the PHRs securely. Furthermore, the system uses a reputation system to track an oracle's behavior and give a rating score to identify the misbehaving oracles. Thus, the system lets them fetch, store securely, and share medical data. Zhang et al [22] presented the architecture for sharing clinical data based on blockchain. The architecture used the FHIR standard and blockchain to solve clinical data interoperability and is called FHIRChain. The FHIRChain helps enable collaborative clinical decision-making among physicians. It also allows for the sharing of clinical data in a trustless and decentralized environment and for auditing through the smart contract. Shahnaz et al [23] designed the role-based access control (RBAC) framework for EMRs using smart contracts. They focused on solving the scalability problem of blockchain via the off-chain scaling mechanism.

Tanwar et al [24] proposed a permission-based system architecture that could share eHealth data using blockchain. In this architecture, patients can join the blockchain network through the client application and update their eHealth data on the blockchain network via chain code. They can also grant or revoke permission to clinicians and researchers for their eHealth data. In conclusion, the architecture achieves patient-centric eHealth data sharing. Figueroa et al [25] used attribute-based access control for the security of a radio frequency identification system for health care. They focused on solving system problems such as scalability, synchronization, and single point of failure using blockchain. Ultimately, the system offers access control to use the medical assets from a suitable location. Daraghmi et al [26] designed a blockchain-based EMR management system called MedChain. They improved the block time and system performance using proof of authority. They also used time-based smart contracts for the privacy and monitoring of EMRs. In brief, they provided a secure environment, data integrity, auditability, and accessibility using authentication techniques, hash function, and proxy re-encryption. Kaur et al [27] proposed blockchain-based storage for securely sharing and querying eHealth data. The storage uses CouchDB considering the unstructured eHealth data. It also stores EMRs in the off-chain and hash of EMRs on the blockchain to ensure the integrity of EMRs and improve the efficiency of storage. Guo et al [28] proposed the multi-authority ABE scheme for cloud-based telemedicine systems. In particular, the scheme protects the integrity of eHealth data (eg, diagnostic opinions) using the blockchain. Furthermore, the scheme updates and revokes the access policy easily.

Most studies [20-28] only focused on blockchain for secure sharing and ensuring the integrity of eHealth data among hospitals. They generally mentioned traceability and accountability via the blockchain, but they did not represent methods for monitoring and accountability. However, these methods should be presented to ensure a secure environment. In particular, accountability is essential in a health care research platform in open environments. For these reasons, unlike other studies, the HBDP focused on the description of the detection method based on the blockchain to ensure accountability. In addition, in the HBDP, even if eHealth data are exported, the data are not ensured usability as they can only be decrypted and used in the HBDP. As mentioned previously, this study is the first to focus on accountability to use eHealth data in an open environment securely.

Table 1. Strengths and weaknesses of blockchain-based studies and the health care big data platform (HBDP).

Studies	System name	Security solutions	Strength	Weakness
Yang et al [20]	—[a]	• Encryption • Access control	• Interoperability among existing health care systems	• The proposed architecture only focused on reading health records and did not discuss sharing of health records.
Madine et al [21]	—	• Encryption • Blockchain	• A patient-centric PHR[b] management system is proposed.	• Not useful for an emergency where the patient is not able to delegate permission
Zhang et al [22]	FHIRChain	• Audit • Access control	• No SPoF[c] problem and fine-grained access control	• The architecture only presented the possibility of health data tracking.
Shahnaz et al [23]	—	• Access control	• The proposed architecture solves the scalability problem of blockchain via off-chain scaling.	• The proposed architecture requires transaction costs and fees for access control.
Tanwar et al [24]	—	• Access control	• A patient-centric eHealth data sharing is achieved.	• Lack of flexible and fine-grained access control
Figueroa et al [25]	—	• Access control	• No SPoF problem and fine-grained access control	• The architecture requires transaction costs and fees for access control.
Daraghmi et al [26]	MedChain	• Encryption • Authentication	• Efficient consensus mechanism and ensuring privacy via time-based smart contracts	• No detailed description of the implementation of the proposed system using the PoA[d]
Kaur et al [27]	—	• Authorization	• Sharing of unstructured eHealth data and off-chain storage	• Not useful for an emergency where the patient is not able to delegate permission
Guo et al [28]	—	• Encryption • Access control	• The ABE[e] scheme is proposed as suitable for the distributed telemedicine system.	• The specific method is not presented to ensure the traceability of the schema.
Ours	HBDP	• Encryption • Audit • Access control	• The detailed methods for accountability in access control are proposed.	• The platform focuses only on 3 SRs[f].

[a]Not presented.

[b]PHR: personal health record.

[c]SPoF: single point of failure.

[d]PoA: proof of authority.

[e]ABE: attribute-based encryption.

[f]SR: security requirement.

Health Care Research Platforms

The use of eHealth data in health care research can fundamentally improve health care owing to the rapid development of big data analytical technologies. For this reason, several studies have proposed health care research platforms that can be used for research using eHealth data. In this section, we review the literature with a focus on the security perspective of these research platforms. Ozaydin et al [29] proposed the design of a data warehouse, which is a Healthcare Research and

Analytics Data Infrastructure Solution (HRADIS). The HRADIS focuses on infrastructure for integrating disparate eHealth data to improve the efficiency of health care. The HRADIS includes an account management framework for RBAC for some eHealth data. Lunn et al [30] proposed a cloud-based digital health research platform for a national longitudinal cohort study. The platform collects and manages eHealth data of sexual and gender minority adults. In this platform, all microservices are within the subnet using virtual private cloud, and eHealth data at rest are stored in the MySQL DB securely after encryption. Furthermore, the platform uses open authorization for programming interfaces, SMS text message–based 2-factor authentication, and logging services to identify malicious users and ensure the security of eHealth data. Ashfaq et al [31] described the regional health care information platform in Halland, Sweden. The platform basically operates within Swedish regulations and the GDPR regarding patient data. On the platform, eHealth data can only be accessed through internal clients secured in the regional IT firewalls. The client can only use related researchers in the approved health care project via the ethical review board in Sweden. In particular, the platform provides anonymized eHealth data to ensure privacy. Conde et al [32] presented an open source–based research platform to support clinical and translational studies, ITCBio. The ITCBio platform supports role and access management tools to promote research collaboration and ensure security. It also provides dynamic consent, which enables ongoing and flexible communication between patients and researchers. De Moor et al [33] described a scalable and adaptable platform for the interoperability of eHealth data systems and clinical research systems. They also presented the security architecture based on many security-related standards in detail. In particular, this architecture supports various security solutions such as identity management and credential delegation. Jones et al [34] proposed the Secure Anonymised Information Linkage databank, which is ensured physical, technical, and procedural control. The Secure Anonymised Information Linkage databank provides encrypted communication and prevents eHealth data from being transferred outside the user's devices. It also performs user authentication via user credentials and 2-factor authentication tokens.

Several studies [29-34] have proposed health care research platforms for using eHealth data. However, most studies have focused on an efficient research environment. Some studies also did not describe security solutions in detail despite the security and privacy of eHealth data being major considerations in health care research platforms. Moreover, as mentioned previously in the Background section, eHealth data subjects are still concerned about the security and privacy of eHealth data. Thus, a study is necessary for a more secure platform for health care research that guarantees the usability of eHealth data while focusing on its security and privacy. The next section proposes a secure and expandable collaborative research platform for health care called the HBDP.

Methods

Overview

This study designed a secure and open environment for health care research. To accomplish this, we first identify potential security threats on a health care research platform. Second, we propose security requirements (SRs) for a secure health care research platform based on these threats. Finally, we present a secure collaborative research platform for health care called the HBDP that can provide a secure analysis environment while meeting these requirements.

Security Threats and Requirements on a Health Care Research Platform

Overview

A health care research platform should properly understand and mitigate security threats to provide a secure analytical environment. This subsection first identifies potential security threats of health care research platforms. We then define the SRs for mitigating these threats.

Security Threats

Various security threats, such as the abuse and illegal export of eHealth data, can arise on a health care research platform. However, we identified well-known security threats in the health care domain as threats to the health care research platform. In other words, many threats can occur on the platform, but we explicitly focused on threats that can occur frequently. A detailed description of the leading security threats is outlined in Textbox 2.

Textbox 2. Leading security threats on a health care research platform.

Leading security threats

- Unauthenticated users: on a health care research platform, unauthenticated users attempt an attack to obtain the authenticated user's credentials [35-37]. In addition, attackers can invalidate the authentication factor to access eHealth data [38]. Hence, a health care research platform must ensure, through user authentication, that only authenticated users have access.
- Unauthorized users: a health care research platform must ensure that only approved eHealth data are available to authorized users through appropriate authorization mechanisms [38,39]. Moreover, the abuse and illegal sharing of eHealth data can occur on a health care research platform even by authorized users. Therefore, a health care research platform also requires a security solution that audits for these activities.
- Leaks of eHealth data by insiders: the greatest security threat for a health care research platform is a breach of eHealth data by insiders [35-37,40]. A prime example of an insider is the eHealth data administrator of the health care research platform. The administrator can easily leak eHealth data as they have general authorization over them. Furthermore, insiders are difficult to detect as they are defined as suitable users within the health care research platform. For these reasons, even if eHealth data on a platform are illegally leaked, the utility of leaked data must not be ensured.

SRs for Mitigating These Threats

A collaborative health care research platform in an open environment should satisfy the diverse SRs that mitigate many types of security threats. However, in this study, we only focused on 3 SRs, which are highly related to accountability in access control for a secure health care research platform based on the aforementioned identified threats. The detailed descriptions of the SRs are outlined in Textbox 3.

Textbox 3. Security requirements (SRs) for mitigating security threats.

SRs for threat mitigation

- SR 1 (access control): access control is a framework that includes authentication and authorization, which is the primary SR and the most important consideration for a health care research platform. It must be performed on this platform so that only authenticated and authorized users can use eHealth data via appropriate devices. For this reason, many existing health care research platforms provide authentication or authorization using various methods [29-34].
- SR 2 (encryption of stored eHealth data): on a health care research platform, the encryption of stored eHealth data ensures the security and privacy of eHealth data when the data are not being used [30,34]. In addition, even if eHealth data are leaked, the data should not be useful. Hence, the encryption of stored eHealth data is one of the most important SRs.
- SR 3 (accountability): when the authenticated and authorized user exports or uses eHealth data via the research platform for health care, the platform administrator or eHealth data provider needs to be able to track and search all the user's activities on the platform at any time. In addition, the platform administrator must identify illegal users and conduct the appropriate follow-up or monitoring in the event of security issues. For these reasons, some health care research platforms provide logging systems or services [30,34].
- Other SRs: the collaborative research platform in an open environment should satisfy various other SRs. For example, anonymization and deidentification are needed for the privacy of eHealth data as the data are sensitive and private [31,32,34]. Secure communication is also necessary to prevent sniffing and tampering with eHealth data and network packets [30,31,34]. In addition, more SRs for the integrity and availability of eHealth data are required [41]. However, as mentioned previously, we focused on the three SRs (ie, access control, encryption of stored eHealth data, and accountability) to support accountability in access control.

Proposed HBDP

Overview

The HBDP uses ABE for the privacy and access control of eHealth data. In particular, the privacy of eHealth data is ensured through column-level encryption even if insiders leak the data. The platform also uses a smart contract to record user activities (eg, log-in and decryption) in the blockchain. Thus, the blockchain allows platform administrators to identify illegal users and conduct appropriate follow-up and monitoring. In other words, the blockchain operates as a distributed logging system in real time and ensures the integrity and nonrepudiation of recorded user activities. In this section, to present the HBDP, we first explain the assumptions and main components in a framework. We then describe the phases of the HBDP in detail.

Assumptions

To describe a framework and scenarios of HBDP, we first define some assumptions. In particular, we present assumptions about other SRs (eg, secure communication, deidentification, integrity, and availability) that the HBDP does not cover. The detailed assumptions are outlined in Textbox 4.

Textbox 4. Assumptions about other security requirements.

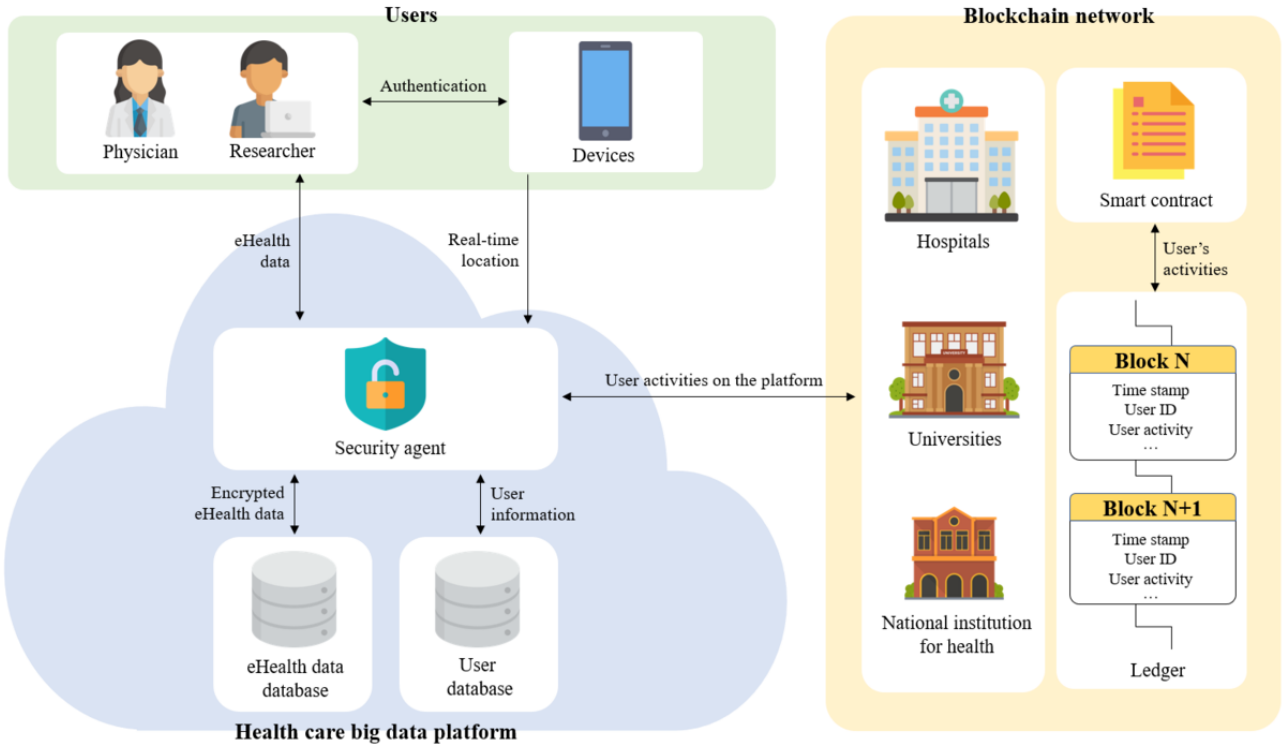
Assumptions

- As eHealth data contain a considerable amount of personally identifiable information, they are generally provided to users after deidentification and anonymization on the platform. Data linkage is also performed to increase the usability of eHealth data before they are provided to users. However, this study did not cover deidentification, anonymization, and data linkage. Thus, all the eHealth data on the platform are assumed to be deidentified and linked via trusted third-party organizations. In addition, eHealth data are assumed to be provided by institutions registered on the platform.
- This study did not cover secure communication between the health care big data platform (HBDP) and the users. Therefore, we assume that the HBDP is securely communicating with its users by using transport layer security protocol–based communication, which is used for secure communication on the internet and across networks. This assumption also holds for communication on the blockchain network.
- Users are assumed to be researchers or physicians with a specific institution that is registered on the HBDP. Thus, they do not need to prove that they are researchers or physicians affiliated with the institution when they register on the platform, but authentication and authorization for access to eHealth data for users are performed on the platform. Furthermore, we also assume that the HBDP provides a variety of analytical tools and methods for researchers to efficiently analyze eHealth data and that the user analysis process is recorded on the distributed ledger, including the analytical tools used.

Main Components

We present a secure collaborative platform for health care research that ensures the privacy and security of eHealth data, called the HBDP. Figure 2 shows a brief overview of our proposed framework for the HBDP. Our proposed framework has 3 main components: users, the HBDP, and the blockchain network. A detailed description of the main components is outlined in Textbox 5.

Figure 2. Overview of our proposed framework for the health care big data platform.



Textbox 5. Main components of our proposed framework.

Framework components

- Users: physicians and researchers who analyze and use eHealth data to treat patients or use them for health care research are representative of this group. They should be required to have a device such as a smartphone or a fingerprint scanner with a GPS for authentication and access control on the platform.
- Health care big data platform (HBDP): the HBDP keeps eHealth data secure and provides an environment where users can use and analyze the data. The platform consists of a security agent and databases (DBs) in a cloud computing environment. DBs are configured as eHealth DBs and user DBs. The eHealth data DB stores eHealth data. The user DB stores user information such as the user ID, hashed password, and user attributes (eg, user department and position). The security agent is a key component of the platform. It performs encryption and decryption of eHealth data using attribute-based encryption. It also requests, as a blockchain client, the blockchain network to record or obtain user activities.
- Blockchain network: the blockchain network consists of a single smart contract, a distributed ledger, and peers. The transactions recorded on the blockchain network are immutable unless the ledgers of all peers are modified. For this reason, the blockchain can be used as a distributed logging system that provides strong accountability, so we use the blockchain network for the tracking of user activities on the HBDP. More specifically, the blockchain communicates with the security agents on the HBDP and helps ensure the accountability and nonrepudiation of the platform. Peers are health facilities and research institutes registered on the platform. They can be endorsing peers or committing peers depending on their system performance. The smart contracts record user activities with time stamps on the distributed ledger, which helps the distributed ledger in the blockchain act as logs for the HBDP.

Phases of the HBDP

Overview

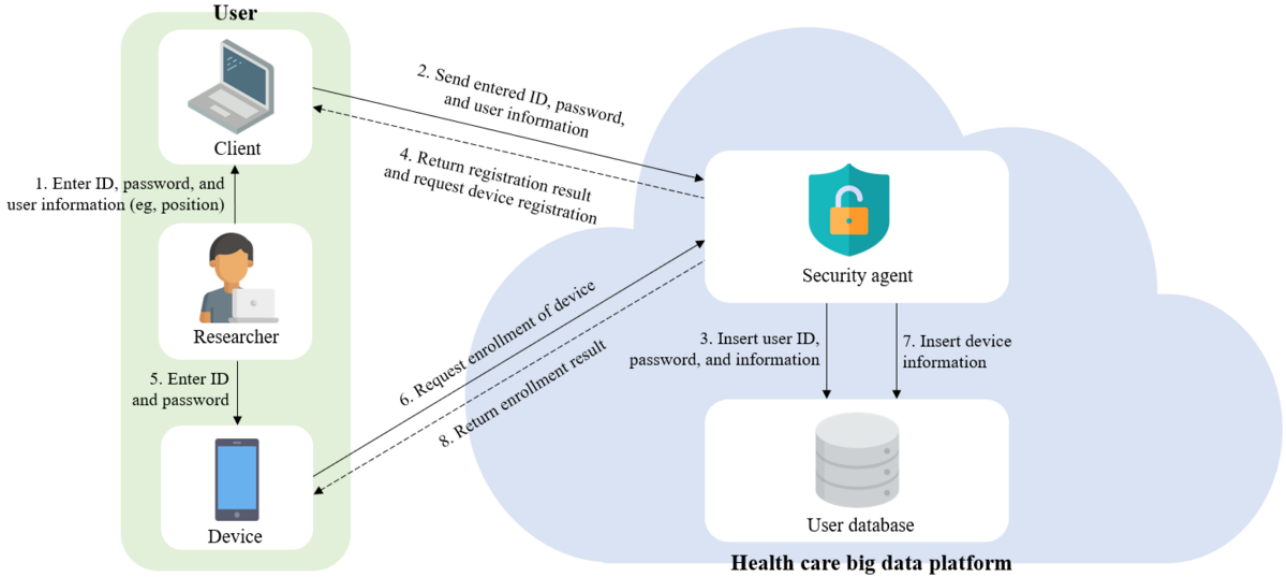
To support secure analytical environments, the HBDP has 4 phases (ie, user registration, storage, download, and use). Each phase is configured to satisfy our defined SRs (ie, the user registration and download phases meet the access control requirement, the storage phase meets the encryption of stored eHealth data requirement, and the use phase achieves accountability). A detailed description of each phase is provided in the following sections.

User Registration Phase

The user registration phase is the first operation for authentication in "access control," which is one of the SRs of a health care research platform. This phase is stored with the user ID and attributes in the user DB on the HBDP. Figure 3 shows a sequence of the user registration phase; the details are described in this section. The user accesses the HBDP and enters the user ID, password, and attributes (eg, the user's department and position). At this time, we assume that this user is authorized by an institution participating in the HBDP. The security agent on this platform then inserts the entered user information into the user DB. The security agent requests device enrollment from the user with the registration result. The user accesses the platform using their device and enters the registered ID, password, and device identifier. The user device then requests the security agent to enroll it along with the entered ID and password. The security agent performs password-based authentication using the received ID and password. If this

authentication is successful, the device ID value is inserted into the user DB, and the security agent relays the result of the device enrollment to the device. After that, the user can access the HBDP at any time.

Figure 3. User registration phase.

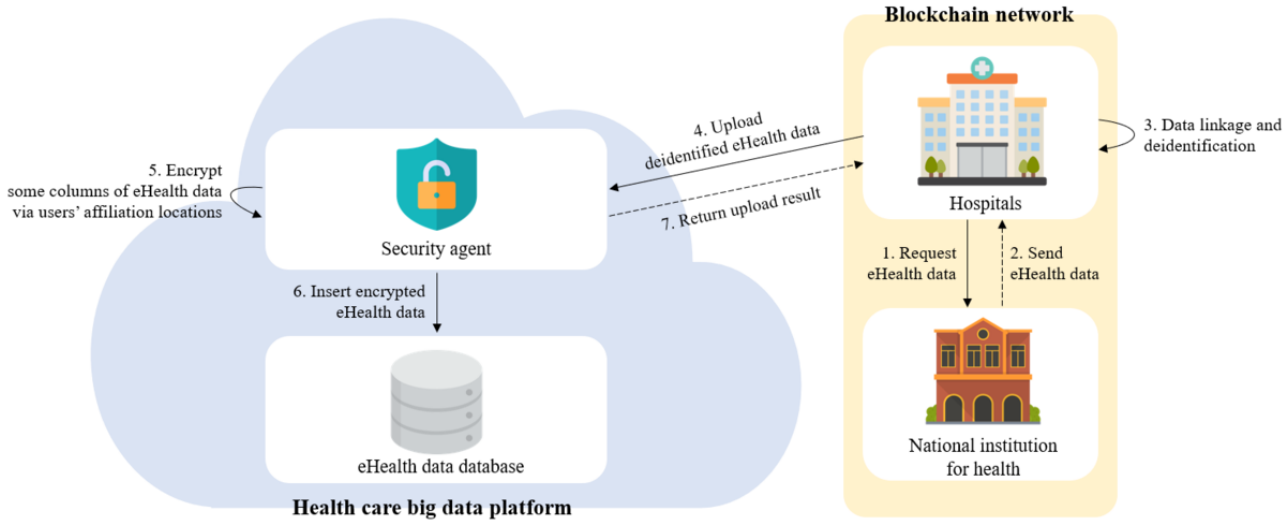


Storage Phase

Attempts to leak eHealth data stored by the HBDP may frequently occur. Even if eHealth data are inevitably leaked in these attempts, the nonusability of the data should be ensured. For this reason, in the storage phase, as shown in Figure 4, the security agent encrypts eHealth data using the locations of institutions registered on the platform before storing the eHealth data on the platform.

In particular, some columns are sensitive columns that provide usability for researchers during the analysis of eHealth data or can be combined with other big data to identify individuals. The storage phase is the operation that ensures the "encryption of stored eHealth data," which is one of the SRs for health care research platforms. This phase ensures that, even if eHealth data are illegally shared or leaked by the administrator or a malicious attacker on the platform, their usability is not ensured because of column-level encryption. Moreover, column-level encryption allows users to use decrypted eHealth data only at their institutions as the decryption of eHealth data fails if the user's real-time location does not match the user's institution.

Figure 4. Storage phase.



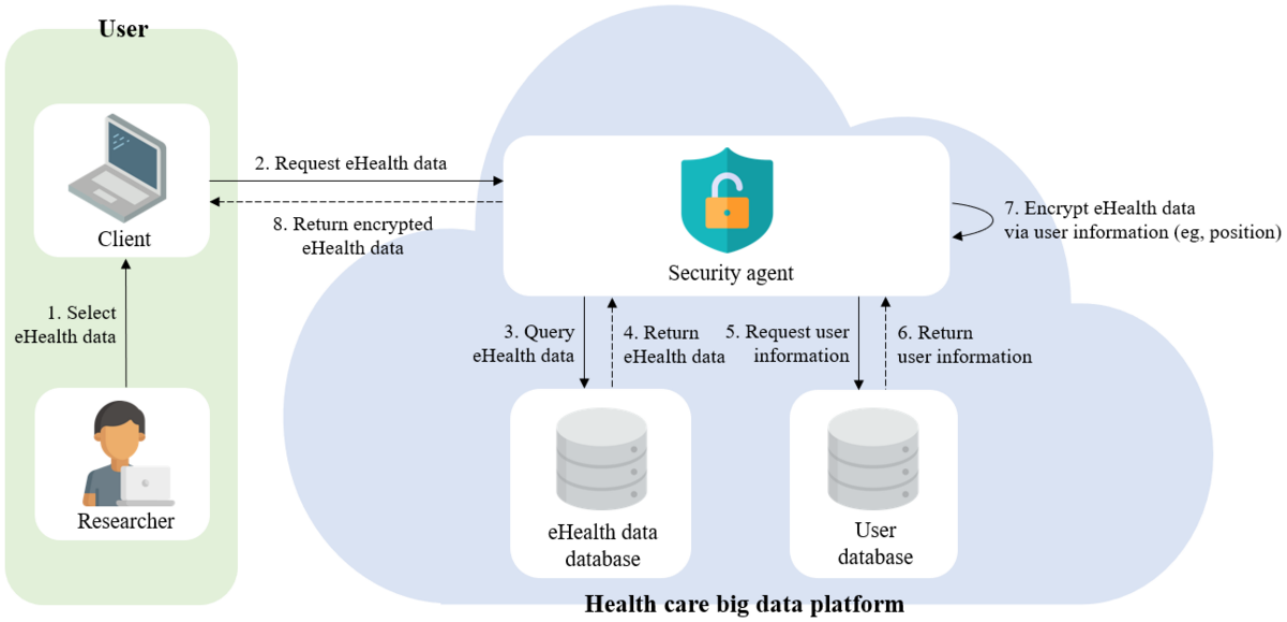
Download Phase

The download phase is the prerequisite for the authorization process in the "access control" for the SRs of the health care research platform. This phase encrypts all contents of column-level–encrypted eHealth data via user attributes (eg, user position and department) to enable the user to download eHealth data. Therefore, the data provided during the download phase are encrypted, so they are impossible to analyze even if a third party obtains them. The download phase is not an essential phase but can increase the efficiency when collaborative research is conducted. For example, if all

collaborative researchers are authorized on the platform, the researcher sends eHealth data after the first analysis to another researcher for collaboration. Another researcher can then proceed with further work based on the analyzed eHealth data on the platform. Figure 5 shows the download phase on the platform. A detailed description of the download phase is provided in the following paragraph.

The user first logs into the HBDP and selects the eHealth data on the download page. The security agent on the platform then sends a query for the eHealth data requested by the user to the eHealth data DB. The eHealth data DB provides column-level–encrypted eHealth data to the security agent. After that, the security agent requests user information, such as the user's attributes and ID, from the user DB. The user DB provides the requested user information to the security agent. The security agent encrypts the column-level–encrypted eHealth data one more time using the provided user information and offers the full encrypted eHealth data. The eHealth data provided during the download phase are encrypted, so they are impossible to analyze even if a third party obtains them.

Figure 5. Download phase.



Use Phase

The use phase performs "access control," which is one of the SRs for the health care research platform. This phase also ensures "accountability" as the use activity of the user on the platform is recorded with a time stamp in the distributed ledger. Figure 6 shows the sequence of the use phase; the details are described in the following paragraphs.

The user logs into the HBDP (ie, password-based authentication) and uploads encrypted eHealth data. The security agent on the platform then requests user and device information from the user DB for decryption and fingerprint authentication. The user DB provides the security agent with the user attributes and user device ID. The security agent uses the device ID value to request fingerprint authentication and real-time location from the user device. The user enters a fingerprint via the enrolled device on the platform for second user authentication. If the user succeeds in fingerprint authentication, the device sends a real-time GPS location to the security agent. The security agent decrypts the eHealth data uploaded by the user using the location and user attributes. In particular, at this time, full decryption is performed on encrypted rows, and column-level decryption is then performed on column-level–encrypted columns (ie, sensitive columns). After the decryption of the eHealth data, the security agent requests the blockchain peer to record the decryption result with a time stamp in the blockchain. Blockchain peers execute a smart contract to record the decryption result on the distributed ledger, and the smart contract inserts the decryption result on the distributed ledger. When the decryption result is recorded in the distributed ledger, the smart contract returns the execution result to the blockchain peers. The blockchain peers send the received execution result to the security agent. Finally, the security agent provides the decrypted eHealth data to the user, and the user is able to use the data only on the platform. In conclusion, for the user to use eHealth data on the HBDP, the data must be decrypted on the platform.

Algorithms 1 and 2 are a pseudocode of the security agent. Algorithms are composed of password-based authentication and decryption of the eHealth data along with the detection of illegal users. In particular, algorithm 1 (Figure 7) first checks whether the user's ID is locked; if the user's ID is not locked, it authenticates the user's credentials. The security agent then requests to record the result of the user log-in activity in the blockchain.

By contrast, algorithm 2 (Figure 8) first performs fingerprint-based authentication for the user's real-time location. After that, the security agent decrypts the data using the created policy via user information, including the location. Finally, the security agent requests to record the result of the use activity of the user in the blockchain. In particular, before recording the result from the procedure, the security agent checks the distributed ledger for the most recent consecutive failed activities in the previous records.

Figure 6. Use phase.

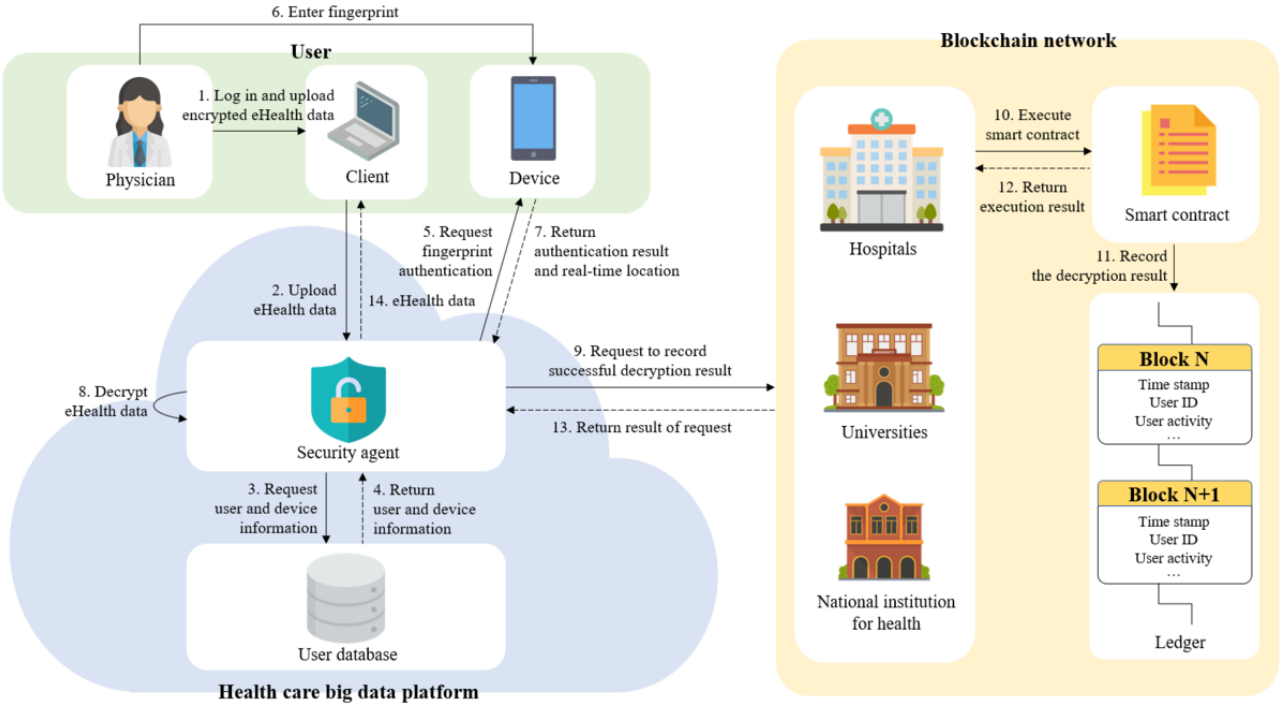


Figure 7. Algorithm 1: pseudocode of authentication for the security agent.

```
Input: user_id, user_password, user_device_id

procedure authentication (user_id, user_password, user_device_id):
    /* block_point is where the user is blocked. */
    Query block_point to the user database using user_id;
    if block_point is "login" then
        Request fingerprint authentication to the user device;
        Send email about login try with a timestamp to the platform admin;
        return loginPage ();
    else if block_point is "utilization" then
        Send an email about decryption try with a timestamp to the platform admin;
        Print "Contract to the platform admin. Your ID is locked !";
        return loginPage ();
    else
        Authenticate the user using user_id and user_password;
        if authentication is successful then
            Record successed login activity with a timestamp to blockchain network;
            return mainPage ("authenticated user");
        else
            Request last activity about login to blockchain network with user_id;
            if more than two last login activity is failed then
                Update block_point to "login" in the user database using user_id;
                Block the user and then send an e-mail about failed login with a timestamp to
                the platform admin;
            end if
            Record failed login activity with a timestamp to blockchain network;
            return loginPage ();
    end if
end procedure
```

Figure 8. Algorithm 2: pseudocode of use for the security agent.

```
Input: enc_eHealth_data, user_attributes (eg, position and department), user_device_id

procedure utilization (enc_eHealth_data, user_attributes, user_device_id):
    fingerprintAuth (user_device_id);
    if fingerprint authentication is failed then
        Print "fingerprint authentication is failed...";
        return user_uploadPage ();
    end if
    Add real_time_location into user_attributes;
    Create policy using user_attributes;
    for each row in enc_eHealth_data do
        Decrypt row using policy;
    end for
    if decryption is failed then
        Request last activity about utilization to blockchain network with user_id;
        if the last utilization activity is failed then
            Update block_point to "utilization" in the user database using user_id;
            Block the user and then send an e-mail about failed decryption with a timestamp to
            the platform admin;
            Print "Contract to the platform admin. Your ID is locked !";
        return loginPage ();
        end if
        Record failed decryption activity with a timestamp to blockchain network;
        return user_uploadPage ();
    end if
    Record successed decryption activity with a timestamp to blockchain network;
    return viewPage (decrypted_eHealth_data);
end procedure
```

Implementation

To demonstrate and prove the feasibility of the HBDP, we implemented the main components of a framework based on the software development life cycle. The software development life cycle is generally configured as requirement analysis, design, implementation, testing, and evolution steps. Following these steps, we first identified SRs (see the SRs in the Security Threats and Requirements on a Health Care Research Platform section). Second, the components were designed based on 3 identified SRs (ie, access control, encryption of stored eHealth data, and accountability). Third, we implemented these components. Textbox 6 shows the specifications for the configuration and implementation environment in detail. We configured the blockchain network for detecting illegal users and a cloud environment to create a scalable, collaborative, and secure environment in the HBDP. In particular, we built the cloud environment using OpenStack (Open Infrastructure Foundation), an open-source cloud operating system, and then developed a web server using the Python-based Flask framework (Python Software Foundation) [42]. We also developed an Android app for user authentication, the security agent for detecting illegal users and monitoring, and a chain code to record and manage user activities. Fourth, the components are tested using a security analysis of the HBDP via case studies in the Results section. Finally, the components are analyzed in the Discussion section to evaluate them.

Figure 9 shows an overview of our implementation and the interactions between the main components. As mentioned previously, our implementation is a proof of concept for demonstrating the features realized by the HBDP.

On a research platform for health care, the cloud environment not only provides various big data analytical tools in the form of software as a service but also provides an environment where researchers can collaborate. The cloud environment also provides scalability and an open environment.

Figure 10 shows some pages from the HBDP. Figure 10A is a page that is shown when a user successfully logs in by entering a registered user ID and password. Users who successfully log in can access this platform at any time and download eHealth data that can be used on the page shown in Figure 10B.

Textbox 6. Specifications of our implementation environment.

Environment specifications

- Processor: Intel Xeon processor E5620 2.40 GHz
- Memory: 32 GB
- Operating system: Ubuntu Linux 18.04.5 LTS
- Smartphone: Galaxy S21 (SM-G991N)
- Languages: Go language, Java, Python, and C++
- Docker engine: version 20.10.7
- OpenStack: version 5.2.0 (Stein)
- MySQL: version 5.7.36
- Android: version 11
- Hyperledger Fabric: version 1.4

Figure 9. Overview of our implementation.

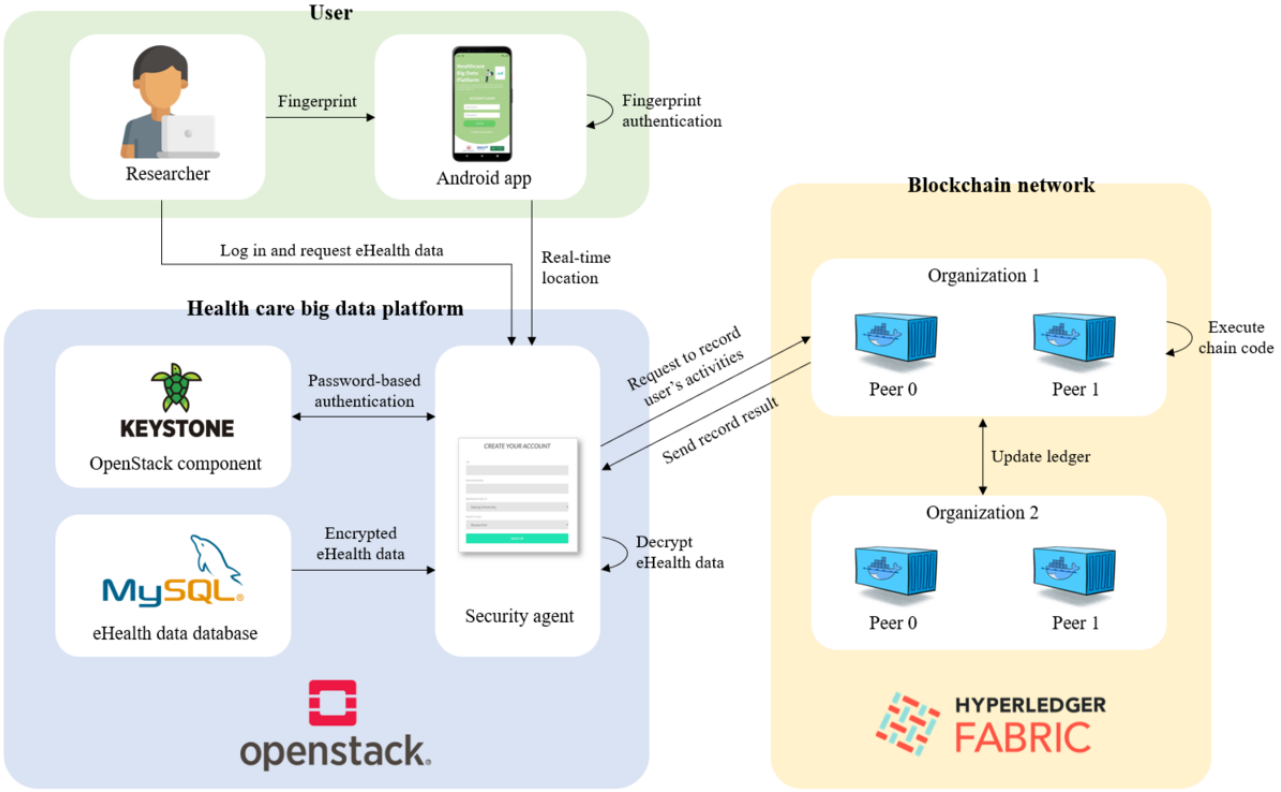


Figure 10. Implemented health care big data platform (HBDP).

(A) Main page

(B) Download page

We also developed an Android app to use fingerprint authentication and real-time location information. In particular, we used Firebase (Firebase Inc) [43] for messaging the Android app. Figure 11 shows some pages of the implemented app. In Figure 11A, the app performs user registration and device enrollment on the platform. In addition, Figure 11B shows device enrollment for fingerprint authentication and real-time location information. Finally, Figure 11C shows a fingerprint authentication request when the user wants to use encrypted eHealth data.

Furthermore, we used the OpenABE library (Zeutro) to perform the encryption and decryption of eHealth data. The eHealth data were anonymous and deidentified open data from the Korea Disease Control and Prevention Agency. The data were from the National Health and Nutrition Examination Survey in South Korea, and they included ID, gender, age, region, and income. In particular, we defined the sensitive columns in our implementation as age and region. Figure 12 shows the eHealth data at each phase on the platform. Figure 12A shows the eHealth data with column-level encryption when they are stored on the platform by the security agent. In addition, when the user downloads eHealth data, they are offered after full encryption of the column-level–encrypted eHealth data using the user attributes stored on the platform, as shown in Figure 12B. Finally, Figure 12C shows the decrypted eHealth data, which can be used by authorized users on the platform.

Generally, blocks in the blockchain are identified by hashes, and the blocks are connected because they have a hash of the previous block. In other words, alteration is impossible unless the blocks of all participants are modified. Therefore, on the platform, the private blockchain is used as a decentralized persistent logging system. We built the blockchain network using Hyperledger Fabric (The Linux Foundation) and designed a smart contract for accountability. Figure 13 shows a web page for detecting illegal users (ie, unauthenticated and unauthorized users) using the distributed ledger. This page helps platform administrators search for specific user activities as well as identify and respond to the actors when security threats arise. In short, the distributed ledger in our implementation provides accountability and nonrepudiation to the HBDP.

Figure 11. Android app.

(A) Main page (B) Enrollment page (C) Authentication page

Figure 12. Encryption and decryption of eHealth data in the health care big data platform.

ERSON_ID	SEX	AGE	REGION
10019903	2	AAADJaETqm/Jqk0SSN	AAADJaETqm/J+6s+fOiCsMr
10164465	1	AAADJaETqm/JxpXvGC	AAADJaETqm/JNsNmbDfppS
10208628	2	AAADJaETqm/JRXx8nn	AAADJaETqm/JLh5WAHCGX
10244958	1	AAADJaETqm/Jrtn+qJF	AAADJaETqm/J6/VZPYOyvHl
10249150	2	AAADJaETqm/J2c+YBe	AAADJaETqm/JxZBxm3Sems
10310174	1	AAADJaETqm/JXSwpT	AAADJaETqm/J5y03EcphC87
10328653	2	AAADJaETqm/J0InkIbC	AAADJaETqm/JPwQYysI3N9A
10334446	1	AAADJaETqm/J9olckvS	AAADJaETqm/JNI1I3O8K7jnc
10368620	2	AAADJaETqm/J9OZRI0	AAADJaETqm/J3nj8goEe+2n
10165731	1	AAADJaETqm/JjgGWXI	AAADJaETqm/J6Zm1XfO+xk
10173339	1	AAADJaETqm/Jf1hDZE	AAADJaETqm/JT1p8a2xa+iD
10222078	1	AAADJaETqm/JcRH/qC	AAADJaETqm/JjF5pn2SEkALI

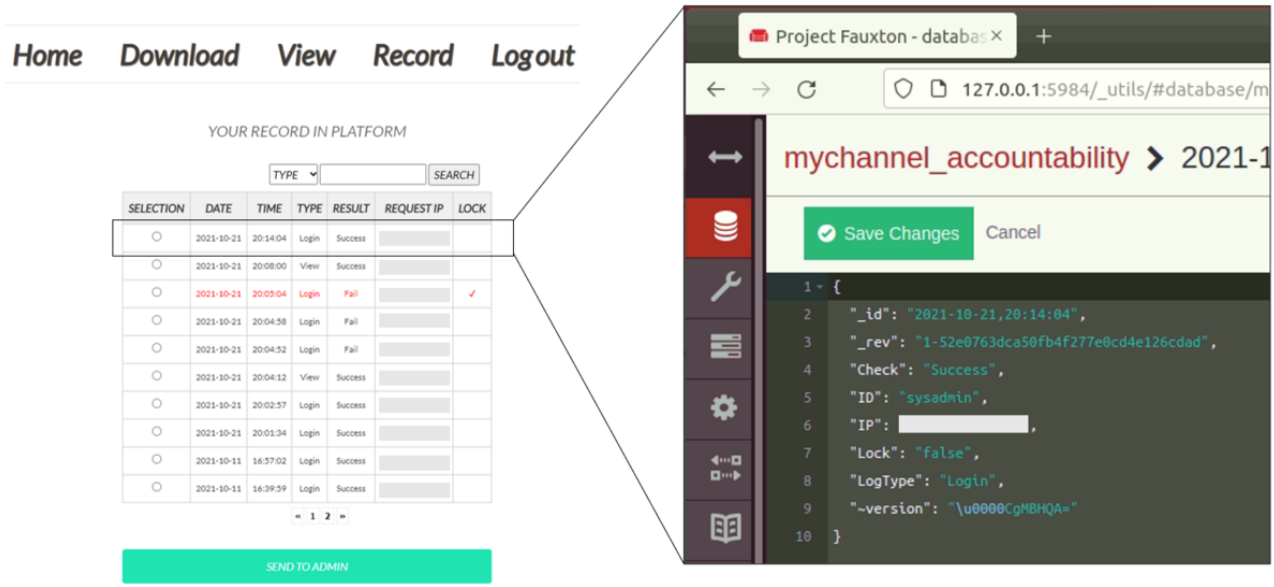
(A) Column-level encrypted eHealth data

(B) Fully encrypted eHealth data

PERSON_ID	SEX	AGE	REGION	SIDO	INCM5
10000670	1	44	1	11	3
10002702	2	53	1	6	3
10004312	1	38	1	1	5
10005116	1	41	1	8	3
10007644	2	50	1	11	2
10011042	2	80	1	1	4
10013000	1	60	1	8	5
10013152	1	28	1	8	2

(C) Fully decrypted eHealth data

Figure 13. Web page for detection of illegal users.



Results

Overview

To show the proof of concept, the previous section implemented the HBDP. This section describes case studies of illegal user detection for security analysis via the implemented HBDP. This section also presents the results of several conducted experiments, which show the efficiency of the private blockchain and ABE cryptography.

Case Studies of Detection of Illegal Users

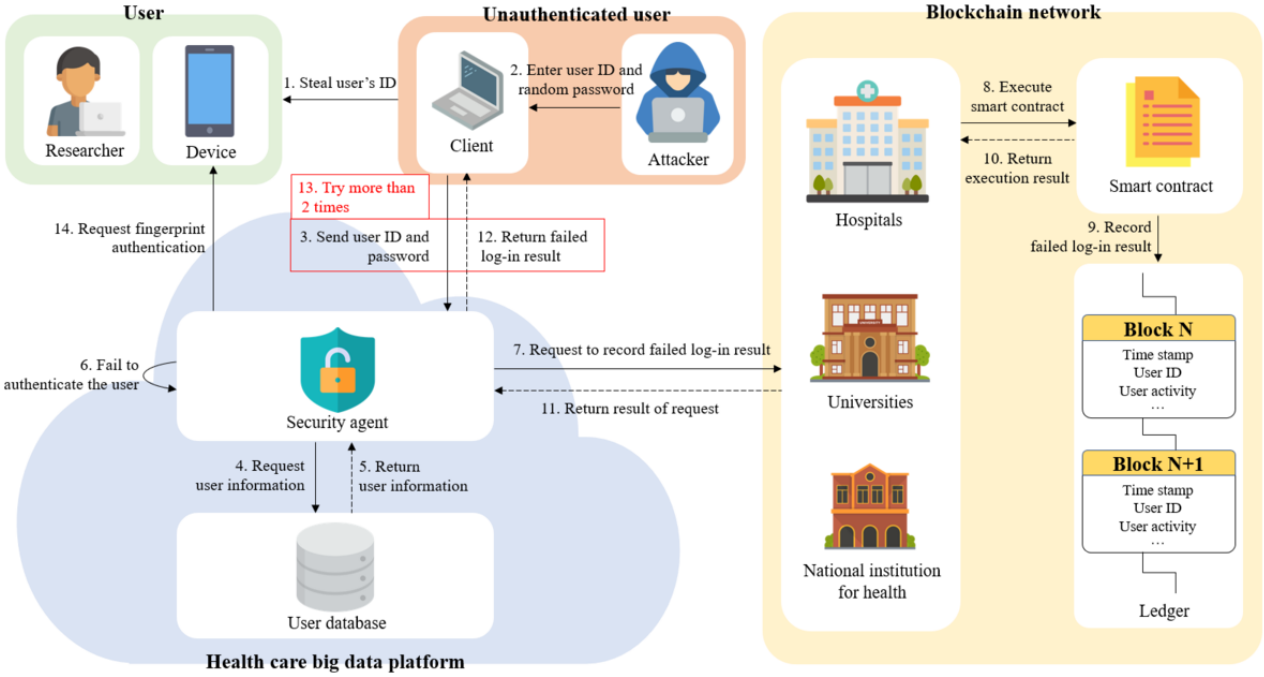
Overview

In addition to the aforementioned security threats, many security threats (eg, the misuse and abuse of eHealth data) can arise on a health care research platform in an open environment. Thus, a secure health care research platform must be able to detect and trace the threats. This subsection describes how to detect two representative security threats—unauthenticated and unauthorized users—through the distributed ledger on the HBDP in an open environment. Moreover, we present the possibility of detecting other security threats through monitoring and access control processes.

Detection of an Unauthenticated User

One of the most common attacks attempted by unauthenticated users is the brute-force attack. Therefore, in this scenario, we assume that an unauthenticated user continuously tries to log into (ie, launches a brute-force attack on) the HBDP by stealing the user ID. Figure 14 shows the sequences for detecting an unauthenticated user on the platform. A detailed explanation of this case study is provided in the following paragraphs.

Figure 14. Detection of unauthenticated user.

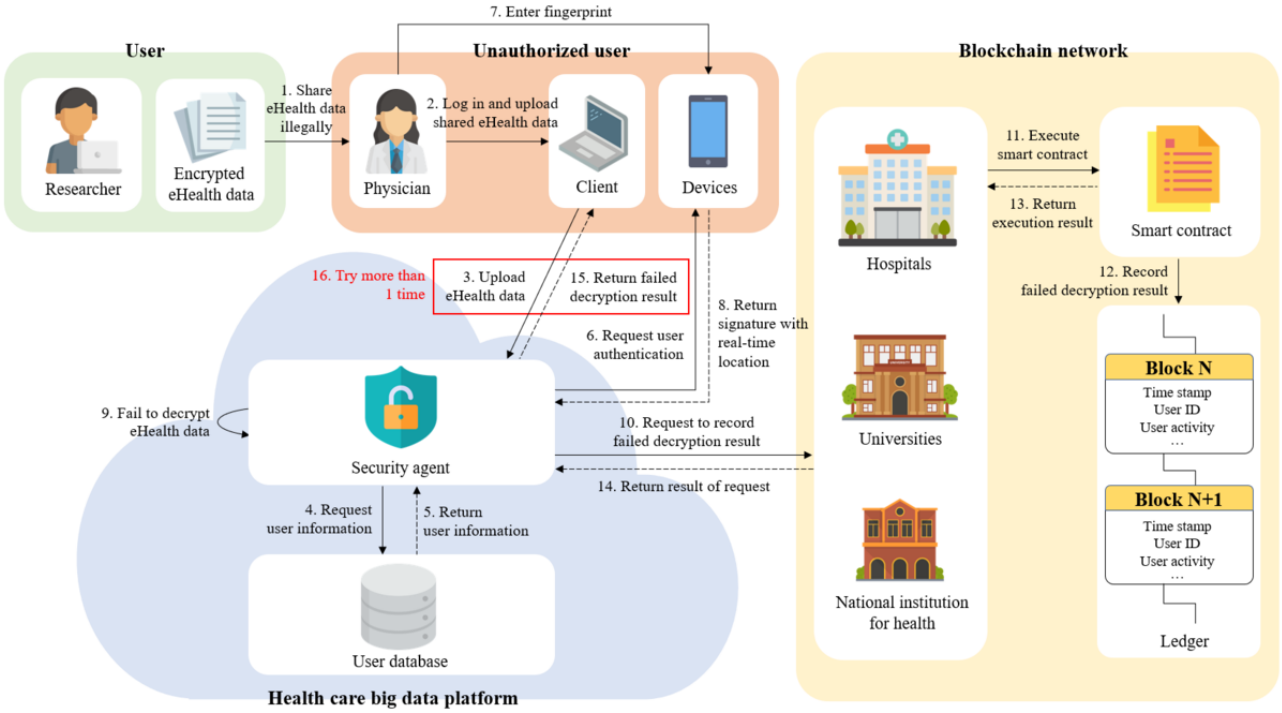


The attacker first steals the user ID that is used on the HBDP and then connects to the platform and inputs the stolen user ID and a random password. The security agent that receives the user ID and password retrieves the user information from the user DB and performs password-based user authentication. After that, the security agent requests the blockchain peer to record the log-in result in the distributed ledger. The blockchain peer records the log-in result through the execution of a smart contract and sends the recorded result to the security agent. The security agent receives the recorded result and then informs the attacker of the log-in failure. The attacker receives the result of the failed log-in and then continuously tries to log in using a random password with the stolen user ID. If the aforementioned sequence is repeated and the log-in fails 2 more times, the user ID is blocked by the security agent. In addition, the security agent requests fingerprint authentication to unblock the user ID from the device enrolled in the HBDP. As a result, platform users will be able to recognize that there has been an illegal log-in attempt. Furthermore, the security agent forwards these attempts to the platform administrator to help them analyze illegal log-in attempts based on the distributed ledger in detail.

Prevention of the Misuse (or Abuse) of eHealth Data by an Unauthorized User

The eHealth data that can be downloaded from the HBDP depend on the user's department and position. For this reason, there is a possibility that unauthorized users can receive encrypted eHealth data from an authorized user. Therefore, we assume that an unauthorized user of the eHealth data wants to use illegally shared or leaked eHealth data. Figure 15 shows the procedure for detecting the illegal sharing of eHealth data. The details are described in the following paragraphs.

Figure 15. Detection of unauthorized user.



The unauthorized user first obtains encrypted eHealth data in the wrong way from the authorized user. After that, the unauthorized user connects to the HBDP and uploads the illegally shared eHealth data to be used after log-in to the platform. The security agent that receives the eHealth data obtains user information from the user DB and requests fingerprint authentication from the enrolled device. The unauthorized user performs fingerprint authentication. If authentication is successful, the device sends a real-time location along with the signature to the security agent. The security agent then decrypts the eHealth data with the received attributes and real-time location after verification of the signature. However, the decryption of eHealth data fails because the unauthorized user's attributes do not match the user attributes used for the encryption of the eHealth data. The security agent requests that the decryption result be recorded on the distributed ledger and then informs the unauthorized user that the decryption has failed after the unauthorized user's activity is recorded on the ledger. If the aforementioned sequence is repeated and the decryption fails one more time, the user ID is blocked by the security agent. The security agent also sends these attempts to the platform administrator to help them analyze illegal use attempts based on the ledger in detail.

In our implementation, various security threats can be detected and blocked, as can unauthenticated and unauthorized users. For example, even if attackers try to decrypt the eHealth data by stealing the user's ID and password, decryption is impossible as fingerprint authentication fails. In addition, even if fingerprint authentication is successful by manipulating the device, decryption is impossible because of incorrect real-time location information. In other words, the HBDP has ensured the security and privacy of eHealth data. Furthermore, the platform administrator can detect illegal users through periodical monitoring as all user activities on the HBDP are recorded in the distributed ledger. In particular, to use even leaked data, they should be decrypted on the HBDP depending on the use

phase. Thus, the administrator can detect this behavior through monitoring. Finally, the HBDP also does not ensure the usefulness of eHealth data via column-level encryption even if leaks by malicious users occur. In conclusion, the HBDP can provide researchers with an open and secure environment in which to efficiently analyze eHealth data.

Performance Evaluation of the HBDP

Overview

Our main concepts are the proposal of a secure research platform for health care and the detection of illegal users using the distributed ledger. For this concept, we presented case studies in the previous section. However, performance is an important factor in proving system efficiency, so we briefly present and describe a performance evaluation of the implemented HBDP in this section.

Average Time for Cryptography

To measure the average time, we performed 10 rounds of encryption and decryption with changes in the number of rows and sensitive columns, as shown in Textbox 7.

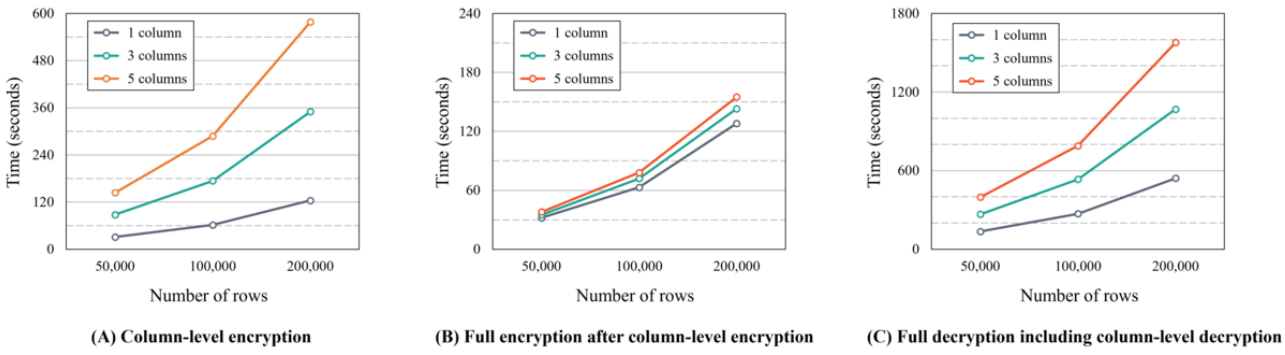
Figure 16 shows the average encryption and decryption times for the number of rows per number of sensitive columns. Figure 16A shows the average column-level encryption time based on the number of rows. In Figure 16A, if the maximum number of rows is 200,000 and the number of sensitive columns is 5, the maximum average time is approximately 10 minutes. Furthermore, Figure 16B shows the average full encryption time for changes in the number of rows versus each number of encrypted sensitive columns when the user downloads eHealth data from the platform (see the Download Phase section). Significantly, as this work encrypts the rows, the number of encrypted sensitive columns does not greatly affect the full encryption time. In other words, the encryption time is not dramatically increased with an increase in the number of encrypted sensitive columns. Figure 16C shows the average full decryption including column-level decryption time for the number of rows versus each number of encrypted sensitive columns when the user uses eHealth data on the HBDP (see the Use Phase section). As this work performs decryption twice, the decryption time required increases dramatically compared with full encryption. If the maximum number of rows is 200,000 and the number of sensitive columns is 5, the average time is approximately 27 minutes, so the HBDP has limitations in use for actual cases. However, we believe that several methods can solve this problem. A discussion of these methods is detailed in the following subsections.

Textbox 7. Simulation parameters for evaluation of cryptography.

Simulation parameters

- Rounds: 10
- Number of rows: 50,000, 100,000, and 200,000
- Number of sensitive columns within 46 columns: 1, 3, and 5
- Type of cryptography: column-level encryption, full encryption after column-level encryption in the download phase, and full decryption including column-level decryption in the use phase

Figure 16. Average time of cryptography.



(A) Column-level encryption (B) Full encryption after column-level encryption (C) Full decryption including column-level decryption

Blockchain Performances

The private blockchain is a distributed logging system that helps detect illegal users on the HBDP. For this reason, we did not evaluate the block and query times and focused only on the accountability and nonrepudiation provided by blockchain features. Therefore, in this section, the write and read times of the designed smart contract are evaluated using Hyperledger Caliper (The Linux Foundation) [44]. We first executed 5 rounds of writing transactions onto the ledger of the blockchain network, with 1000 transactions in each round at rates of 100, 150, 200, 250, and 300 transactions per second (TPS), as shown in Textbox 8. We then executed 5 rounds of reading transactions into the ledger's blockchain network at rates of 100, 150, 200, 250, and 300 TPS, with 1000 transactions in each round after writing 100 transactions. In particular, at this time, we assume that the platform administrator searches 100 records of previous user activity.

Figure 17 and Figure 18 show the average latencies and throughputs of our executions. In Figure 17A, the 1Org with 1Peer in write mode takes <3 seconds in 300 TPS, which is a

much lower latency than other networks. Conversely, the 1Org with 1Peer in write mode has a higher throughput (approximately 150 TPS) than other networks, as shown in Figure 18A. The 1Org with 1Peer in read mode has an average latency of approximately 14 seconds and a throughput of approximately 63 TPS in 300 TPS, as shown in Figure 17B and Figure 18B. For the 2Orgs with 2Peers in read mode, the average latency and throughput reach approximately 19 seconds and 47 TPS, respectively, in 300 TPS. The results show that many organizations and peers reached high latency and low throughput in both read and write modes, so the latency and throughput are inversely proportional in write mode.

Textbox 8. Simulation parameters for evaluation of the blockchain.

Simulation parameters

- Rounds: 5
- Transactions: 1000 each
- Transaction rates: 100, 150, 200, 250, and 300
- Transaction mode: read and write
- Networks: 2Orgs with 2Peers, 2Orgs with 1Peer, and 1Org with 1Peer
- Orderer: solo

Figure 17. Average transaction latencies.

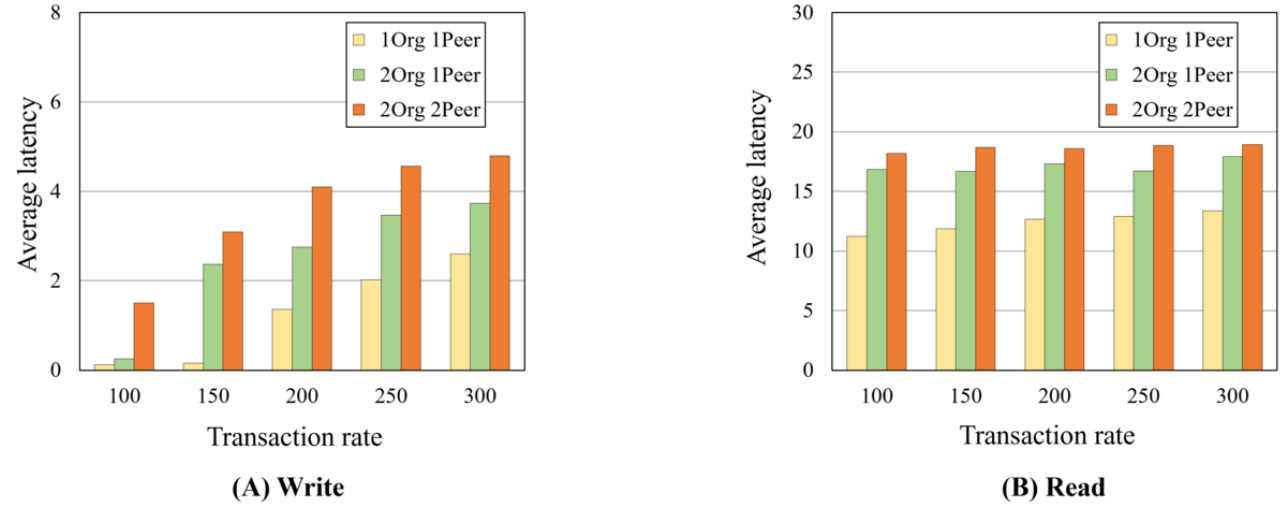
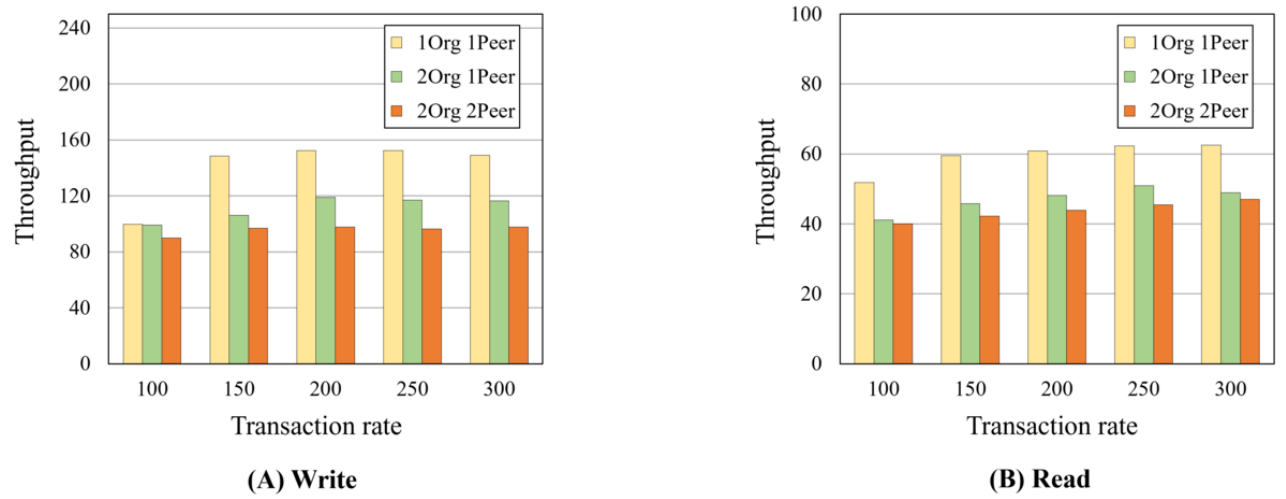


Figure 18. Transaction throughputs.



Discussion

Principal Findings

In the Results section, we first conducted a security analysis of the HBDP. The results showed that the HBDP provides a secure environment. We then presented the average times for cryptography and blockchain performance of the HBDP to evaluate its efficiency. As a result, some performances (eg, high decryption time depending on the number of encrypted columns) showed to need improvement. Therefore, we first discuss these results in this section. Some limitations of the HBDP are then presented. In addition, we compare the HBDP with prior works. Finally, we describe our further works on blockchain and IoT.

The HBDP provides a secure research environment but has several challenges to solve to be efficient. Thus, this subsection discusses our results and these challenges in detail. First, to prove accountability in the HBDP, we presented case studies on detecting unauthenticated and unauthorized users via the implemented HBDP. In addition, we described some methods to detect other security threats via the process of access control and monitoring. The results showed that the HBDP supports accountability in access control.

Second, as the number of sensitive columns increased, column-level encryption and full decryption including column-level decryption times increased significantly in our results. This issue would cause inconvenience to users in a real environment. To address this issue, we present some solutions. First, when eHealth data have many sensitive columns, efficient cryptography times can be achieved by merging these columns into 1 column to perform encryption and decryption. Second, the security agent can be configured first to perform decryption of some columns to show eHealth data and then decrypt other columns in the background process. This method may make the user feel that the delay is minimal compared with the previous approach. Finally, in the cloud environment, multiple security agents can be configured to perform parallel processing [45-47]. This method is efficient and the most widely used approach. Unfortunately, our work did not use these methods, but they are expected to provide better cryptography times.

Finally, the read mode in blockchain has higher latency and lower throughput than the write mode by approximately ≥60%. In general, blockchain performance has higher latency and lower throughput in write mode than in read mode, but our evaluation showed the opposite result. This result might have been due to the process of looking up and reading all the records of 100 transactions. Nevertheless, the blockchain can be used sufficiently as a distributed logging system for the HBDP as it did not show poor performance. To prove this effectiveness, our discussion can be extended through a performance comparison evaluation with existing studies. However, we did not conduct the performance comparison because of differences in the implementation environment and configuration. The blockchain performance is generally affected by the role configuration of peers and orderers, differences in consensus algorithms, and blockchain type. Furthermore, even if a smart contract performs the same function, the performance can vary depending on how the smart contract is implemented. In conclusion, we consider that a comparison of the performances is useless in perfect nonequivalent environments.

Limitations

A health care research platform must offer an efficient environment as the primary purpose of the HBDP is to analyze eHealth data and then use the resulting values for research. For this environment, the interoperability of eHealth data, analytical and visualization tools, and data linkage are needed, but the HBDP implemented a few functions for eHealth data. Hence, in our future work, the HBDP will offer an efficient research platform that provides various analytics and visualization tools (eg, Hadoop, Tableau, and Spark) as software as a service.

Furthermore, in this study, the HBDP only focused on three SRs (ie, access control, encryption of stored eHealth data, and accountability), but various additional SRs (eg, deidentification) are needed for a more secure environment. In addition, even after deidentification of eHealth data, the possibility of reidentification remains. Thus, our future work should also provide other SRs and methods to reduce the risk of reidentification through reidentification assessments in advance [48,49].

Finally, the main scope of this study was access control and accountability for a research platform for health care, so the HBDP did not ensure the integrity and availability of eHealth data. In the HBDP, by also writing the hash of eHealth data on the distributed ledger, integrity can be ensured but not completely. For this reason, the HBDP needs solutions to ensure the integrity and availability of eHealth data for a complete research platform for health care. Availability and integrity can be generally ensured by existing cryptography technologies (eg, diverse types of firewalls, message authentication codes, intrusion detection systems, and hash functions). We also consider that some studies [50-54] are helpful for an efficient health care research platform.

Comparison With Prior Work

For discussions of the HBDP, this section compares the HBDP with the existing health care research platforms based on defined SRs. Table 2 shows that the HBDP and previous research platforms met specific SRs. First, access control (ie, authentication and authorization methods) was partially addressed in all studies [29-34] and was also addressed in the HBDP. In particular, some of the studies [31,32,34] granted access to eHealth data through direct approval from the relevant authorities or contracts. However, this access control is cumbersome and complex, and there is a possibility of overly limiting the use of data. Some studies [29,34] used RBAC for access control, but RBAC cannot easily provide fine-grained access control. By contrast, ABE encryption generally achieves fine-grained access control as administrators can create detailed security policies using various attributes [55-57]. ABE has also been able to achieve flexible access control recently [58,59]. Hence, the HBDP enables fine-grained access control through the various sensing data of IoT devices as we use ABE encryption, unlike existing research platforms for health care. Furthermore, two studies that provided encryption for stored eHealth data were those by Lunn et al [30], Jones et al [34], and

the HBDP. This encryption is needed to prevent illegal leaks by insiders and malicious attackers. Even anonymized eHealth data must be encrypted when they are stored on a research platform to make reidentification difficult and useless if the eHealth data are leaked. Finally, accountability is an audit trail that helps the platform administrator take appropriate action when a security incident occurs and mitigate security threats via monitoring. However, the logging system was implemented in a centralized form in the studies by Lunn et al [30] and Jones et al [34]. A centralized system has difficulty operating a logging service when the system is unavailable, and there is a possibility that the integrity of logs can be undermined by attackers. The HBDP uses a logging system in a decentralized form via blockchain and, thus, even if 1 node is unavailable, logging is still possible and ensures the integrity of logs as all peers own the distributed ledger. Moreover, we presented detailed methods of illegal user detection to prove accountability in the HBDP, unlike previous platforms.

Table 2. Comparison of the health care big data platform (HBDP) and related studies.

SRs[a]	Studies						
	Ozaydin et al [29]	Lunn et al [30]	Ashfaq et al [31]	Conde et al [32]	De Moor et al [33]	Jones et al [34]	Ours
SR 1[b]							
Authentication		✓		✓	✓	✓	✓
Authorization	✓		✓	✓	✓	✓	✓
Fine-grained access control							✓
SR 2[c]—encryption (encryption level and content of eHealth data)		✓ (Full data and medical report)				✓ (Sensitive data and identifiable data)	✓ (Sensitive data and medical condition)
SR 3[d]							
Decentralization							✓
Centralization		✓				✓	
Illegal user detection methods							✓

[a]SR: security requirement.

[b]Access control.

[c]Encryption of stored eHealth data.

[d]Accountability.

Private Blockchain

eHealth data subjects hope to strengthen their rights by participating directly in eHealth data access decisions. They are also concerned with the privacy and security of eHealth data. However, when they directly participate in access decisions, it has the potential to stifle or unduly limit the usability of eHealth data in research (eg, the approval of researchers' requests to use eHealth data is delayed for a long time or they are unconditionally refused). Therefore, the subjects' rights must be ensured in other ways. With the distributed ledger of the blockchain, we expect that providers can supervise although not directly participate in access decisions. For example, eHealth data subjects can easily search the use history and users of their eHealth data at any time via the distributed ledger on the platform and object to the use if there are any issues. In addition, as the recorded history of the blockchain is difficult to alter, it is expected to elicit greater trust from eHealth subjects. Although we did not implement this supervisory function, we expect that further research will help address concerns about the use of eHealth data as well as advance the rights of eHealth data subjects.

Interoperability on IoT Devices

The eHealth data from various sensors and IoT devices are rapidly increasing and being collected in many health facilities. These eHealth data can improve public health and provide high-quality customized health care services when they are used in research, so they must be offered on a health care research platform. In general, IoT devices are connected to and managed by IoT platforms. However, it is currently difficult to share or use collected eHealth data because of the various interoperability issues on IoT platforms. In particular, secure interoperability cannot be guaranteed as each IoT platform has different access control methods and security policies for IoT devices. Therefore, our future research will provide various and detailed eHealth data to researchers by ensuring the secure interoperability of IoT platforms on the HBDP.

Conclusions

The use of eHealth data in health care research offers promising potential and advantages. However, eHealth data are more sensitive than other big data as they contain more personal information, so the privacy and security of eHealth data must be ensured for them to be used in studies. In addition, eHealth data subjects are still concerned about unauthorized data reuse and sharing within existing health care research platforms. Thus,

we designed a more secure collaborative platform for health care research called the HBDP. This platform ensures the privacy and security of eHealth data using a private blockchain and ABE cryptography. The private blockchain operates as a decentralized persistent log DB in which all activities occurring on the HBDP are recorded with time stamps. As a result, the records in the blockchain (ie, distributed ledger) help platform administrators and users detect and respond to unauthenticated and unauthorized users on the HBDP. ABE cryptography ensures privacy even if eHealth data are leaked from the platform and enables detailed and fine-grained access control using situational information. Furthermore, we developed and tested the HBDP, blockchain network, and an Android app using Hyperledger Fabric, OpenStack, and OpenABE library to show the feasibility of the platform. We also described the detection of illegal users (ie, unauthenticated and unauthorized users) via case studies. As this study focused only on a secure environment for health care research, some future work is needed to provide an efficient and complete research platform. Nevertheless, we believe that the HBDP will provide a sufficiently secure environment for the use of eHealth data in health care research.

Acknowledgments

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (2021R1A2C2012635).

Authors' Contributions

GK wrote this paper and performed the implementation. YGK supervised and coordinated the investigation.

Conflicts of Interest

None declared.

References

1. Health Insurance Portability and Accountability Act. U.S. Department of Health & Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations [accessed 2022-01-10]
2. General data protection regulation. Intersoft Consulting. URL: https://gdpr-info.eu [accessed 2022-01-10]
3. California Consumer Privacy Act. State of California Department of Justice. URL: https://oag.ca.gov/privacy/ccpa [accessed 2022-01-10]
4. Oh S, Seo Y, Lee E, Kim Y. A comprehensive survey on security and privacy for electronic health data. Int J Environ Res Public Health 2021 Sep 14;18(18):9668 [FREE Full text] [doi: 10.3390/ijerph18189668] [Medline: 34574593]
5. O'Keefe CM, Rubin DB. Individual privacy versus public good: protecting confidentiality in health research. Stat Med 2015 Oct 15;34(23):3081-3103. [doi: 10.1002/sim.6543] [Medline: 26045214]
6. The research platform for healthcare in the South Korea. Ministry of Health and Welfare. URL: https://hcdl.mohw.go.kr [accessed 2022-01-10]
7. Data access request service (DARS). National Health Service Digital. URL: https://digital.nhs.uk/services/data-access-request-service-dars [accessed 2022-01-10]
8. Swiss Personalized Health Network (SPHN) homepage. Swiss Personalized Health Network. URL: https://sphn.ch [accessed 2022-07-18]
9. Atkin C, Crosby B, Dunn K, Price G, Marston E, Crawford C, PIONEER Data Hub. Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff. Res Involv Engagem 2021 Jun 14;7(1):40 [FREE Full text] [doi: 10.1186/s40900-021-00281-2] [Medline: 34127076]
10. Koo J, Kang G, Kim Y. Security and privacy in big data life cycle: a survey and open challenges. Sustainability 2020 Dec 17;12(24):10571. [doi: 10.3390/su122410571]
11. Naveed M, Ayday E, Clayton EW, Fellay J, Gunter CA, Hubaux J, et al. Privacy in the genomic era. ACM Comput Surv 2015 Sep;48(1):6 [FREE Full text] [doi: 10.1145/2767007] [Medline: 26640318]
12. Hyperledger fabric. Hyperledger Foundation. URL: https://www.hyperledger.org/use/fabric [accessed 2022-01-10]
13. zeutro / openabe. GitHub. 2021 Jan 28. URL: https://github.com/zeutro/openabe [accessed 2022-01-10]
14. The most widely deployed open source cloud software in the world. Openstack. URL: https://www.openstack.org [accessed 2022-01-10]
15. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 clinical document architecture, release 2. J Am Med Informatics Assoc 2006 Jan 01;13(1):30-39. [doi: 10.1197/jamia.m1888]
16. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; Jun 20-22, 2013; Porto, Portugal.
17. OMOP common data model. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/data-standardization/the-common-data-model [accessed 2022-01-10]

18. Digital imaging and communications in medicine homepage. Digital Imaging and Communications in Medicine. URL: https://www.dicomstandard.org [accessed 2022-01-10]
19. Noumeir R, Renaud B. IHE cross-enterprise document sharing for imaging: interoperability testing software. Source Code Biol Med 2010 Sep 21;5(1):9 [FREE Full text] [doi: 10.1186/1751-0473-5-9] [Medline: 20858241]
20. Yang G, Li C, Marstein KE. A blockchain‐based architecture for securing electronic health record systems. Concurrency Computat Pract Exper 2019 Aug 04;33(14):e5479. [doi: 10.1002/cpe.5479]
21. Madine MM, Battah AA, Yaqoob I, Salah K, Jayaraman R, Al-Hammadi Y, et al. Blockchain for giving patients control over their medical records. IEEE Access 2020;8:193102-193115. [doi: 10.1109/access.2020.3032553]
22. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
23. Shahnaz A, Qamar U, Khalid A. Using blockchain for electronic health records. IEEE Access 2019;7:147782-147795. [doi: 10.1109/access.2019.2946373]
24. Tanwar S, Parekh K, Evans R. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. J Inform Security Applications 2020 Feb;50:102407. [doi: 10.1016/j.jisa.2019.102407]
25. Figueroa S, Añorga J, Arrizabalaga S. An attribute-based access control model in RFID systems based on blockchain decentralized applications for healthcare environments. Computers 2019 Jul 31;8(3):57. [doi: 10.3390/computers8030057]
26. Daraghmi E, Daraghmi Y, Yuan S. MedChain: a design of blockchain-based system for medical records access and permissions management. IEEE Access 2019;7:164595-164613. [doi: 10.1109/access.2019.2952942]
27. Kaur J, Rani R, Kalra N. Blockchain‐based framework for secured storage, sharing, and querying of electronic healthcare records. Concurrency Computat Pract Exper 2021 May 20;33(20):e6369. [doi: 10.1002/cpe.6369]
28. Guo R, Shi H, Zheng D, Jing C, Zhuang C, Wang Z. Flexible and efficient blockchain-based ABE scheme with multi-authority for medical on demand in telemedicine system. IEEE Access 2019;7:88012-88025. [doi: 10.1109/access.2019.2925625]
29. Ozaydin B, Zengul F, Oner N, Feldman SS. Healthcare research and analytics data infrastructure solution: a data warehouse for health services research. J Med Internet Res 2020 Jun 04;22(6):e18579 [FREE Full text] [doi: 10.2196/18579] [Medline: 32496199]
30. Lunn M, Lubensky M, Hunt C, Flentje A, Capriotti M, Sooksaman C, et al. A digital health research platform for community engagement, recruitment, and retention of sexual and gender minority adults in a national longitudinal cohort study--the PRIDE Study. J Am Med Inform Assoc 2019 Aug 01;26(8-9):737-748 [FREE Full text] [doi: 10.1093/jamia/ocz082] [Medline: 31162545]
31. Ashfaq A, Lönn S, Nilsson H, Eriksson JA, Kwatra J, Yasin ZM, et al. Data resource profile: regional healthcare information platform in Halland, Sweden. Int J Epidemiol 2020 Jun 01;49(3):738-79f. [doi: 10.1093/ije/dyz262] [Medline: 31930310]
32. Conde J, Moreno-Conde A, Salas-Fernández S, Parra-Calderón CL. ITCBio, a clinical and translational research platform. AMIA Annu Symp Proc 2019;2019:673-680 [FREE Full text] [Medline: 32308862]
33. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-173 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.006] [Medline: 25463966]
34. Jones KH, Ford DV, Thompson S, Lyons R. A profile of the SAIL databank on the UK secure research platform. Int J Popul Data Sci 2019 Nov 20;4(2):1134 [FREE Full text] [doi: 10.23889/ijpds.v4i2.1134] [Medline: 34095541]
35. Singh A, Chatterjee K. Security and privacy issues of electronic healthcare system: a survey. J Inform Optimization Sci 2020 Feb 02;40(8):1709-1729. [doi: 10.1080/02522667.2019.1703265]
36. Hathaliya JJ, Tanwar S. An exhaustive survey on security and privacy issues in Healthcare 4.0. Comput Commun 2020 Mar;153:311-335. [doi: 10.1016/j.comcom.2020.02.018]
37. Alvarez M. Security trends in the healthcare industry. IBM. 2017 Feb. URL: https://www.ibm.com/downloads/cas/PLWZ76MM [accessed 2022-07-25]
38. Newaz AI, Sikder AK, Rahman MA, Uluagac AS. A survey on security and privacy issues in modern healthcare systems. ACM Trans Comput Healthcare 2021 Jul 31;2(3):1-44. [doi: 10.1145/3453176]
39. Al-Janabi S, Al-Shourbaji I, Shojafar M, Shamshirband S. Survey of main challenges (security and privacy) in wireless body area networks for healthcare applications. Egyptian Informatics J 2017 Jul;18(2):113-122. [doi: 10.1016/j.eij.2016.11.001]
40. Ahmed A, Latif R, Latif S, Abbas H, Khan FA. Malicious insiders attack in IoT based multi-cloud e-healthcare environment: a systematic literature review. Multimed Tools Appl 2018 Jan 20;77(17):21947-21965. [doi: 10.1007/s11042-017-5540-x]
41. Nasiri S, Sadoughi F, Tadayon M, Dehnad A. Security requirements of internet of things-based healthcare system: a survey study. Acta Inform Med 2019 Dec;27(4):253-258 [FREE Full text] [doi: 10.5455/aim.2019.27.253-258] [Medline: 32055092]
42. Flask framework. Pallets Projects. URL: https://flask.palletsprojects.com/en/2.0.x [accessed 2022-01-10]
43. Firebase homepage. Firebase. URL: https://firebase.google.com [accessed 2022-01-10]
44. hyperledger / caliper. GitHub. 2022 Aug 8. URL: https://github.com/hyperledger/caliper [accessed 2022-01-10]

45. Li Y, Zhang Z. Parallel computing: review and perspective. In: Proceedings of the 5th International Conference on Information Science and Control Engineering (ICISCE). 2018 Presented at: 5th International Conference on Information Science and Control Engineering (ICISCE); Jul 20-22, 2018; Zhengzhou, China.
46. Zhang Y, Cao T, Li S, Tian X, Yuan L, Jia H, et al. Parallel processing systems for big data: a survey. Proc IEEE 2016 Nov;104(11):2114-2136. [doi: 10.1109/jproc.2016.2591592]
47. Ageed Z, Mahmood M, Sadeeq M, Abdulrazzaq M, Dino H. Cloud computing resources impacts on heavy-load parallel processing approaches. IOSR J Comput Eng 2020;22(3):30-41.
48. Simon GE, Shortreed SM, Coley RY, Penfold RB, Rossom RC, Waitzfelder BE, et al. Assessing and minimizing re-identification risk in research data derived from health care records. EGEMS (Wash DC) 2019 Mar 29;7(1):6 [FREE Full text] [doi: 10.5334/egems.270] [Medline: 30972355]
49. Branson J, Good N, Chen J, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA policy 0070 and health Canada regulations. Trials 2020 Feb 18;21(1):200 [FREE Full text] [doi: 10.1186/s13063-020-4120-y] [Medline: 32070405]
50. Yuan J, Yu S. Efficient public integrity checking for cloud data sharing with multi-user modification. In: Proceedings of the IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. 2014 Presented at: IEEE INFOCOM 2014 - IEEE Conference on Computer Communications; Apr 27 -May 02, 2014; Toronto, ON, Canada. [doi: 10.1109/infocom.2014.6848154]
51. Zhang Y, Xu C, Liang X, Li H, Mu Y, Zhang X. Efficient public verification of data integrity for cloud storage systems from indistinguishability obfuscation. IEEE Trans Inform Forensic Secur 2017 Mar;12(3):676-688. [doi: 10.1109/tifs.2016.2631951]
52. Wei P, Wang D, Zhao Y, Tyagi SK, Kumar N. Blockchain data-based cloud data integrity protection mechanism. Future Generation Comput Syst 2020 Jan;102:902-911. [doi: 10.1016/j.future.2019.09.028]
53. Nabi M, Toeroe M, Khendek F. Availability in the cloud: state of the art. J Network Comput Appl 2016 Jan;60:54-67. [doi: 10.1016/j.jnca.2015.11.014]
54. Endo PT, Rodrigues M, Gonçalves GE, Kelner J, Sadok DH, Curescu C. High availability in clouds: systematic review and research challenges. J Cloud Comp 2016 Oct 18;5(1):16. [doi: 10.1186/s13677-016-0066-8]
55. Goyal V, Pandey O, Sahai A, Waters B. Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on Computer and communications security. 2006 Presented at: CCS06: 13th ACM Conference on Computer and Communications Security 2006; Oct 30- Nov 3, 2006; Alexandria Virginia USA. [doi: 10.1145/1180405.1180418]
56. Wang G, Liu Q, Wu J. Hierarchical attribute-based encryption for fine-grained access control in cloud storage services. In: Proceedings of the 17th ACM conference on Computer and communications security. 2010 Presented at: CCS '10: 17th ACM Conference on Computer and Communications Security 2010; Oct 4 - 8, 2010; Chicago Illinois USA. [doi: 10.1145/1866307.1866414]
57. Guo C, Zhuang R, Jie Y, Ren Y, Wu T, Choo KR. Fine-grained database field search using attribute-based encryption for e-healthcare clouds. J Med Syst 2016 Nov;40(11):235. [doi: 10.1007/s10916-016-0588-0] [Medline: 27653042]
58. Li J, Yao W, Zhang Y, Qian H, Han J. Flexible and fine-grained attribute-based data storage in cloud computing. IEEE Trans Serv Comput 2017 Sep 1;10(5):785-796. [doi: 10.1109/tsc.2016.2520932]
59. Wan Z, Liu J, Deng RH. HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing. IEEE Trans Inform Forensic Secur 2012 Apr;7(2):743-754. [doi: 10.1109/tifs.2011.2172209]

Abbreviations

ABE: attribute-based encryption
CDA: Clinical Document Architecture
CDM: Common Data Model
DARS: Data Access Request Service
DB: database
DICOM: Digital Imaging and Communications in Medicine
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resources
GDPR: General Data Protection Regulation
HBDP: health care big data platform
HL7: Health Level 7
HRADIS: Healthcare Research and Analytics Data Infrastructure Solution
IoT: Internet of Things
NHS: National Health Service
OMOP: Observational Medical Outcomes Partnership
PHR: personal health record

RBAC: role-based access control
SR: security requirement
TPS: transactions per second
XDS: Cross-Enterprise Document Sharing

Edited by T Leung; submitted 14.03.22; peer-reviewed by P Asprion, L Bošnjak, MDG Pimentel; comments to author 13.07.22; revised version received 02.08.22; accepted 30.08.22; published 14.10.22.

Please cite as:
Kang G, Kim YG
Secure Collaborative Platform for Health Care Research in an Open Environment: Perspective on Accountability in Access Control
J Med Internet Res 2022;24(10):e37978
URL: https://www.jmir.org/2022/10/e37978
doi:10.2196/37978
PMID:36240003

Review

Tracking Openness and Topic Evolution of COVID-19 Publications January 2020-March 2021: Comprehensive Bibliometric and Topic Modeling Analysis

Maider San Torcuato[1], BSc, MD; Núria Bautista-Puig[2,3], BSc, MD, PhD; Olatz Arrizabalaga[1], BSc, MD, PhD; Eva Méndez[2], BSc, MD, PhD

[1]Innovation Unit, Biodonostia Health Research Institute, San Sebastián, Spain

[2]Library and Information Science Department, Universidad Carlos III de Madrid, Getafe, Spain

[3]Library and Information Science Department, Complutense University of Madrid, Madrid, Spain

Corresponding Author:
Olatz Arrizabalaga, BSc, MD, PhD
Innovation Unit
Biodonostia Health Research Institute
Paseo Dr Beguiristain s/n
20014
San Sebastián
Spain
Phone: 34 943006001
Email: olatz.arrizabalaga@biodonostia.org

Abstract

Background: The COVID-19 outbreak highlighted the importance of rapid access to research.

Objective: The aim of this study was to investigate research communication related to COVID-19, the level of openness of papers, and the main topics of research into this disease.

Methods: Open access (OA) uptake (typologies, license use) and the topic evolution of publications were analyzed from the start of the pandemic (January 1, 2020) until the end of a year of widespread lockdown (March 1, 2021).

Results: The sample included 95,605 publications; 94.1% were published in an OA form, 44% of which were published as Bronze OA. Among these OA publications, 42% do not have a license, which can limit the number of citations and thus the impact. Using a topic modeling approach, we found that articles in Hybrid and Green OA publications are more focused on patients and their effects, whereas the strategy to combat the pandemic adopted by different countries was the main topic of articles selecting publication via the Gold OA route.

Conclusions: Although OA scientific production has increased, some weaknesses in OA practice, such as lack of licensing or under-researched topics, still hold back its effective use for further research.

(J Med Internet Res 2022;24(10):e40011) doi:10.2196/40011

KEYWORDS

COVID-19; open access; OA; SARS-CoV-2; scholarly communication; topic modeling; research; dissemination; accessibility; scientometry; publications; communication; research topics

Introduction

Background

On January 30, 2020, the World Health Organization declared the COVID-19 outbreak a “public health emergency of international concern,” and declared a pandemic on March 11, 2020, at which point the virus had infected more than 150,000 people in 154 countries [1-3]. One year later (March 2021) the number of infected people reached 3.8 million worldwide [4].

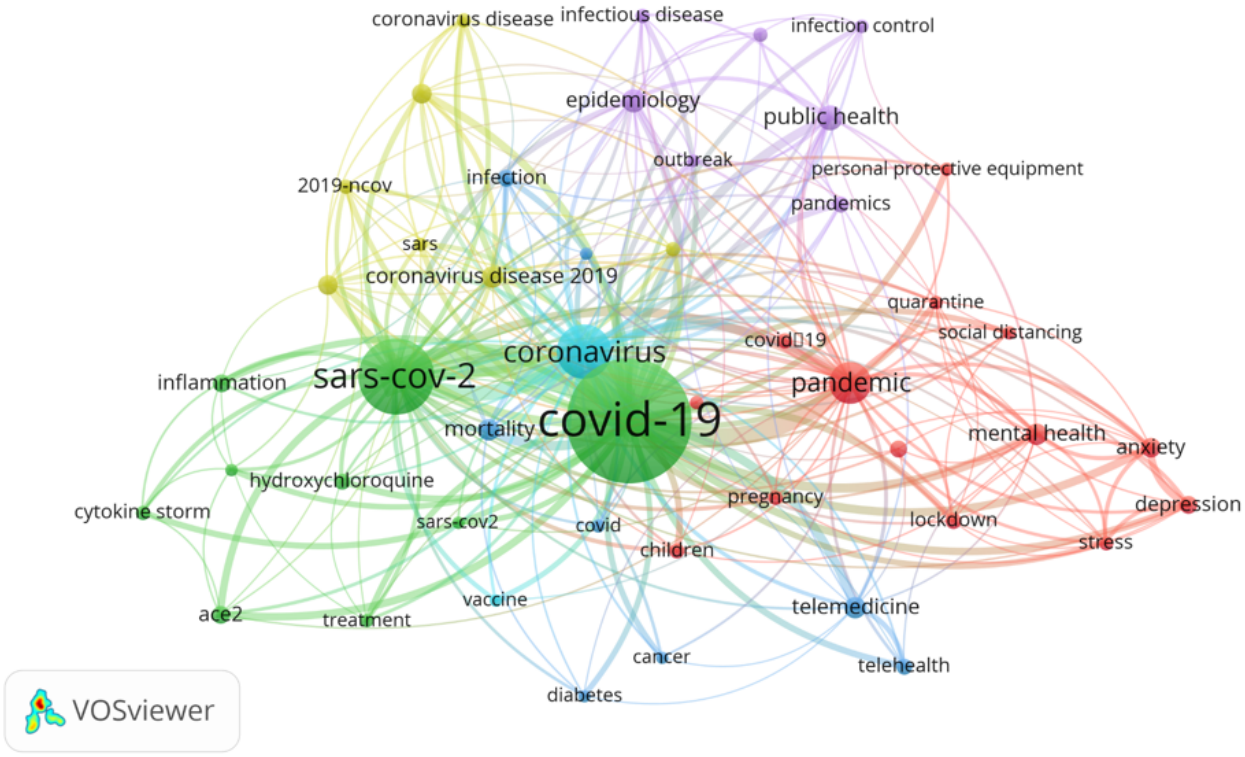
The scientific community is facing one of its greatest challenges for research: to quickly develop solutions for the COVID-19 pandemic. This exceptional situation requires a collective scientific effort that has been reflected daily in the publication of hundreds of scientific documents and resources (ranging from articles and reviews to clinical guides or protocols and data).

We are likely witnessing the greatest concentration ever of scientific resources specifically directed to the resolution of a common problem [5]. The effectiveness of both the publication system and the different components of traditional scientific communication (journals, databases, and repositories) is crucial to perform medical research as well as other types of research focus (ie, economic, educational, psychological) about this new coronavirus, such as delineating risk factors, clinical features, and treatment strategies, including vaccines [6].

Research topics have also rapidly changed during the pandemic, focusing on different areas of interest (Figure 1): COVID-19 and treatment (green cluster), populations at risk (light blue cluster), effects of the pandemic on mental health and impacts of social distancing (red cluster), public health (purple cluster), and coronavirus terms or families (yellow cluster).

We adopted a metaresearch approach to investigate the scholarly communication on this disease, particularly focusing on the open access (OA) uptake, along with the evolution of topics about COVID-19 in different OA publication venues.

Figure 1. Co-occurrence map within the 50 most frequent keywords among cited SARS CoV-2–related publications with at least 200 publications (data extracted from PubMed: January 1, 2020, to March 1, 2021). Image created using VOSviewer [7].



Changes in the Scholarly Publication System

COVID-19 has challenged scientists to overcome the "normal" pace of scholarly communication. The main objection that the current system faced from the beginning of the pandemic is two-fold: science that is closed by default and the overload of articles, with 1000 COVID-19–related publications per week estimated at the beginning of the pandemic in PubMed [5]. As a result, a global health crisis has been readily recognized as an information crisis or "infodemic" [8,9].

During the pandemic, numerous efforts were undertaken to make COVID-19 research publicly available as fast as possible. On January 31, 2020, the Wellcome Trust called on researchers, funders, and journals to share data and make findings immediately available to inform the public health response to this outbreak [10]. Signatories to this statement include relevant publishers (Elsevier, Wiley, Springer, Taylor and Francis, among others). This was also followed by large scientific journals, especially biomedical journals (eg, *JAMA*, *British Medical Journal [BMJ]*, *Science*, Oxford, Cambridge, or *New England Journal of Medicine*) [5], at least temporarily. However, publishers have not always liberated their copyright licenses, and for those who did, it was mainly as an exceptional practice rather than a change of policy.

New pressures and new opportunities were introduced for the scholarly publishing system [11]. Horbach [12] analyzed 669 articles and found that medical journals had accelerated their publication process (eg, the time between submission and publication decreased on average by 49%). However, some studies show evidence of adverse effects, including unethical practices by predatory journals during the pandemic, reduction of journals' quality standards, or biases (eg, most of the scientific output has been from Western countries or English-only publishing at the expense of local communities that could have relevant insights on the topic) [12-14].

State-of-the-Art and Previous Bibliometric Studies

Bibliometric techniques have been used to present an overview of COVID-19 research. Efforts have been made to analyze the coverage of different data sources of COVID-19 publications [15-17], using altmetrics (ie, Wikipedia and Mendeley) [18,19], analyzing the effectiveness and impact of collaboration [20,21],

gender differences [22], topic evolution [16,23], scholarly communication flow during this pandemic [24,25], as well as OA of these research outputs[5,15].

Although a high volume of scientific publications are being produced (150,000 peer-reviewed COVID-19 outputs were published in the Dimensions database between January 2020 and April 2021, and 40,000 COVID-19 preprints were posted in this period), the percentage of publications on OA differs from that of databases, with 72.81% in Dimensions and 88.8% in PubMed [5,11,15,26]. The majority of OA publications follow the "Bronze" route and are mainly published without a license (representing 76.4% of all OA papers recorded at early stages of the pandemic in PubMed) [15]. However, most bibliometric studies and OA analyses were performed in the early stages of the pandemic.

As pointed out by Colavizza et al [16], the early stage of pandemic research was dominated by the topic of the coronavirus outbreak. However, in analyzing 27,370 publications by topics using Medical Subject Heading (MeSH) terms in PubMed, Wang and Hong [23] found that epidemiology and public health interventions have gathered the highest attention. Within these categories, the most popular topics were prevention and control of COVID-19, whereas other topics have been less popular, such as drug therapy. However, little is known about the differences in OA typologies or licenses, which could help researchers and scientific policymakers understand and guide the status of COVID-19 research.

Accordingly, the aim of this study was to investigate the research communication about this disease, the level of openness of papers, and the main topics of research. We also were guided by the following research questions: What effect has the emergency situation had on scholarly communication? How have OA publishing models affected citation rates? What effect does the presence of a proper license have on the citation of published papers? How have the topics covered in the publications evolved during the pandemic? Does the OA publishing model have an effect on the analyzed topics?

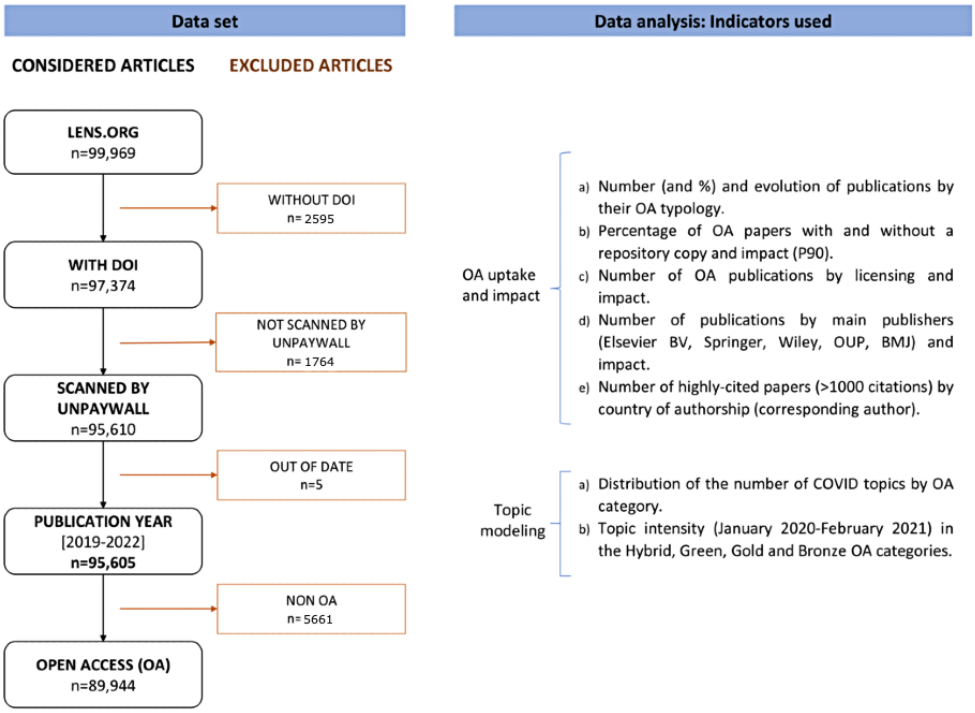
Methods

Sources and Search Strategy

In this study, different databases and tools were used to collect and analyze COVID-19–related publications, relevant information about OA (typology and licenses), and the main topics covered (Figure 2). The platforms chosen were PubMed, Lens, Microsoft Academics, and Unpaywall that collectively cover a large proportion of free biomedical publications. For this study, we selected PubMed as it is the only database that has been able to record the largest number of publications on this topic since the beginning of the pandemic, including early articles, in an updated manner (daily updating). Other databases such as Web of Science (WoS) or Scopus have a delay of indexing relative to PubMed [15,27]. Furthermore, PubMed is a more well-suited database for biomedical research, whereas Scopus and WoS are more multidisciplinary databases. Moreover, PubMed offers free access to all users, while Scopus and WoS are subscription-based.

The search was performed on March 16, 2021, in the Lens data platform (considering only the PubMed database) by the following query, suggested by the National Library of Medicine and the National Center for Biotechnology Information: 2019-nCoV OR 2019nCoV OR COVID-19 OR SARS-CoV-2 OR (wuhan AND coronavirus)

Figure 2. Workflow used to select the sample for the study (sources and indicators).



Data Selection, Scope of the Study, and Limitations

We focused our analysis on the period from January 1, 2020, to March 1, 2021. This period corresponds with the peak cases in the population and the initial vaccination protocol (immunized) [28]. The query retrieved a total of 99,969 scientific works about COVID-19 in PubMed, 2595 (2.60%) of which did not have a DOI and 1764 (1.76%) of which were not recognized by Unpaywall. Therefore, the study considered a total of 95,605 publications. The Lens database was used to collect 1.6 million citations from the selected publications.

Although this represents a very comprehensive study based on the number of publications analyzed and the different types of analyses performed, some limitations must be pointed out. We only considered one database (PubMed), which is mainly skewed toward medical and biomedical publications and does not cover all academic fields nor all publication languages. Non-English publications and nonbiomedical fields are not covered or are under-represented. Another limitation is due to the use of Unpaywall; although this source provides relevant information on OA, it does not have full coverage and sometimes contradicts information in Crossref. Limitations of the search strategy adopted include the use of the keywords for selecting each COVID-19–related article, which conflicts with the contribution of research toward the pandemic and other studies that might presumably include buzzwords.

Data Analysis and Research Steps

We first analyzed the uptake of OA and its impact on scientific publications about COVID-19 during the study period (January 2020-March 2021). Figure 2 summarizes the main indicators analyzed. OA status information was considered because OA aims to maximize access to research by promoting visibility and diffusion of scientific outputs and removing technical or financial barriers [29]. Different OA categories defined by Unpaywall were considered in our analysis: Bronze (articles freely available on websites hosted by their publisher, either immediately or following an embargo, but are not formally licensed for reuse), Gold (articles in fully accessible OA journals by paying a fee, known as an article processing charge [APC]), Green (a copy archived in an online open repository with access to final versions after an embargo period), and Hybrid (articles in a subscription journal made OA by paying the APC). In addition, the total number of citations per article, according to Lens, was considered and analyzed by OA typology. However, considering that a skewed distribution is associated with a risk that the citation statistics are dominated by a few highly cited or uncited papers (eg, published in a short time window), a percentile-based bibliometric indicator is needed. Therefore, in this study, we used the 90th percentile (P90) based on total citations received by each paper, which enabled better cross-OA comparisons of the impact of publications. P90 means that the paper belongs to the top 10% most frequently cited papers, which was calculated using linear interpolation of modes in a spreadsheet.

We also used Unpaywall to collect information about licenses. The main licensing options analyzed were Creative Commons (CC) or publisher-specific licenses. Classified according to their level of reuse, from the most open to the most restrictive, the license types include: American Chemical Society (ACS)-Specific, CC, CC-BY, CC-BY-NC, CC-BY-NC-ND, CC-BY-NC-SA, CC-BY-ND, CC-BY-SA, Elsevier-Specific, Implied-OA, PD, publisher-specific license, and no license. In addition, the publisher information was retrieved by analyzing the five most frequent publishers (Elsevier BV, Wiley, Oxford University Press [OUP], and BMJ). Openrefine was chosen to organize, clean up, and analyze the data. This tool allowed us to filter the data extracted from Lens, connect the data with the Unpaywall application programming interface, and to gather more information about OA and the repositories (PMC or institutional repositories found in Open Archives Initiative-Protocol for Metadata Harvesting [OAI-PMH]). For data analysis, interpretation and visualization of a spreadsheet were also used. We further mapped the country distribution of the corresponding author from 105 highly cited papers (with more than 1000 citations, representing 0.11% of the total) using ArcGIS software.

Next, we applied a topic modeling technique to the titles and abstracts of COVID-19 publications by OA types (Bronze, Gold, Green, and Hybrid) to identify prominent topics during the pandemic and their evolution. This probabilistic technique takes a collection of texts as input and makes it possible to identify and learn “topics” from a corpus of documents [30,31]. The keywords from all documents were then grouped by those that appear closer together (by frequency); thus, it can be argued that they are thematically connected, forming clusters (or topics). As a result of this technique, the biggest cluster in Bronze was composed of keywords such as student, medical, or survey, among others, which constituted cluster 0 (see the full list of clusters in Multimedia Appendix 1).

Unlike clustering, topic modeling assumes that each document will fit into one or more topics. Elimination of stop words, spaces, and other irrelevant characters was performed in R software using the tm package [32,33]. A total of 87,744 papers (87.8%) of the data set were used in this analysis. For topic modeling, we adapted Colavizza et al’s [16] code in Open Jupyter Notebook by training the data set with the latent Dirichlet allocation model using the gensim implementation [16,31,34]. In this case, 15 clusters were defined for the identification of keywords divided by OA type, each composed of a group of keywords (see the full list in Multimedia Appendix 1). To more deeply analyze the content, each cluster was categorized into the main topics defined by Colavizza et al [16] and Wang and Hong [23], as described below. “Coronavirus Outbreaks” and “Epidemics” were merged into a single topic (labeled “Epidemics”) as they included similar clusters. The 5 topics and their scope are defined in Table 1. A comprehensive list of topics and clusters is provided in Multimedia Appendix 1.

In addition to this classification, the monthly topic intensity of the clusters (based on the number of publications) by OA type was analyzed to observe the changes over time. As the period of study covered up to March 1, 2021, March was not included in this analysis.

The data set used in this study has been made available in Zenodo [35].

Table 1. Topic description and examples of identified keywords.

Topic	Definition and scope	Examples of keywords
Clinical Medicine	Study and practice of medicine that is founded on the direct observation of patients	treatment, chest, therapy, symptom, clinical trial
Immunology	Covers the study of immune systems in all organisms	immune, antibody, drug, vaccine, spike
Molecular Biology	Branch of biology dealing with the structure and function of the macromolecules essential to life	proteins, nucleic acids, virus cell, antibodies, cytokine
Public Health	Branch of medicine dealing with public health, including hygiene, epidemiology, and disease prevention	public health system, patient, mental health, community, nursing
Epidemiology	Studies the rapid spread of disease to a large number of people in a given population within a short period of time	disease, outbreak, countries, masks, tests

Results

OA Uptake

Overview

From the 95,605 PubMed articles considered (Figure 2), 98.34% (n=94,015) were journal articles and 94.08% (n=89,944) were published in OA format, with the majority in Bronze OA (44.8%), followed by Gold (31.9%), Green (14.1%), and Hybrid (9.3%) (Figure 3a).

The remaining publications represent posted content (n=1551), book chapters (n=27), and "others" (n=6, including 1 report, 1 peer review, 2 proceeding articles, and 1 uncategorized type) (Figure 3b).

Overall, 41.39% (39,573/95,605) of all publications were published under the Bronze OA model, 29.49% (28,192/95,602) as Gold, 14.64% (13,993/95,605) as Green, and 8.56% (8186/95,605) as Hybrid OA (Figure 3c).

Measuring the P90 of the citation distribution of the field showed that Hybrid, Green, and Bronze OA articles have higher citation values of 29, 26, and 24, respectively, compared to Gold OA articles (16) and articles published in closed journals (5).

Analysis of the evolution of publishing models (Figure 3d) showed that use of the Green model exhibited a decreasing trend during the pandemic, eventually becoming the least-used model. As the pandemic progressed, Bronze and Gold publishing models became more prominent, with a significant increase of the Bronze model from the second quarter of 2020 onward.

Figure 3. PubMed-hosted SARS CoV-2–related papers published from January 1, 2020, to March 1, 2021 and their open access (OA) status based on Unpaywall. (a) Percentage of considered and excluded papers (without DOI and not scanned by Unpaywall) and their OA ratios. (b) PubMed established publication type and their OA type. (c) Percentage of publications and citations divided by their OA publishing model. (d) Evolution of publications according to their OA publishing model. P90: 90th percentile.

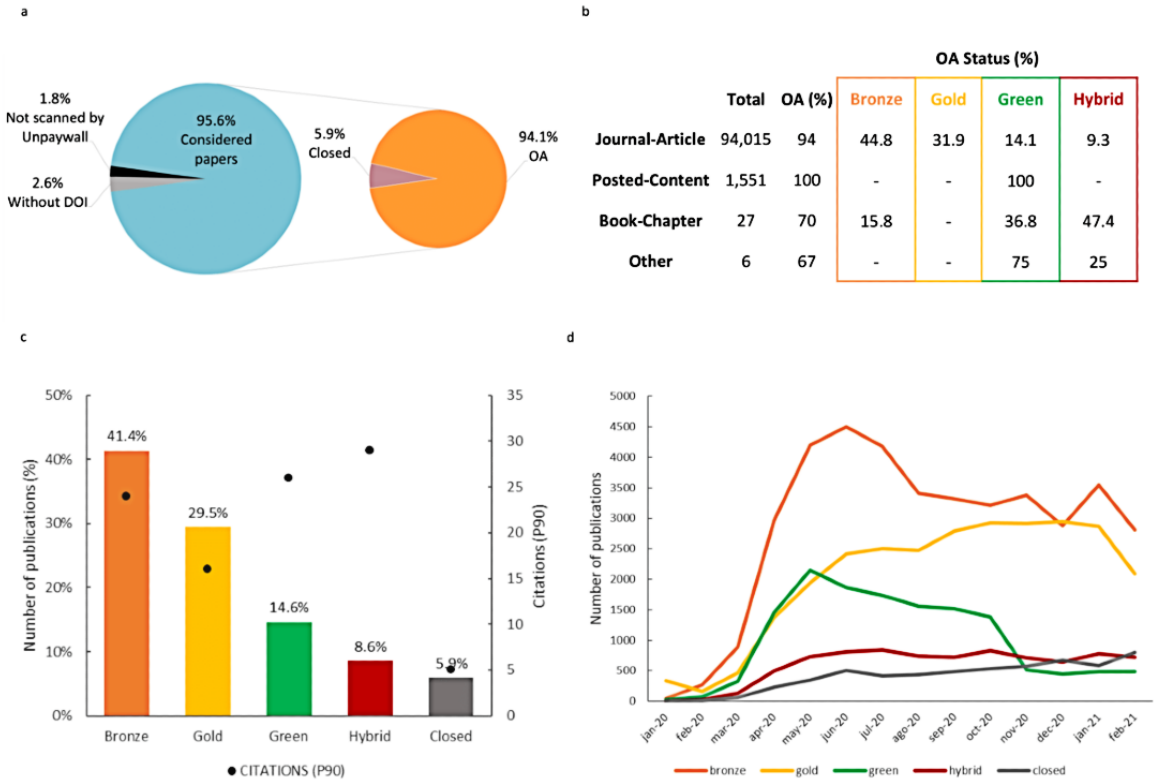
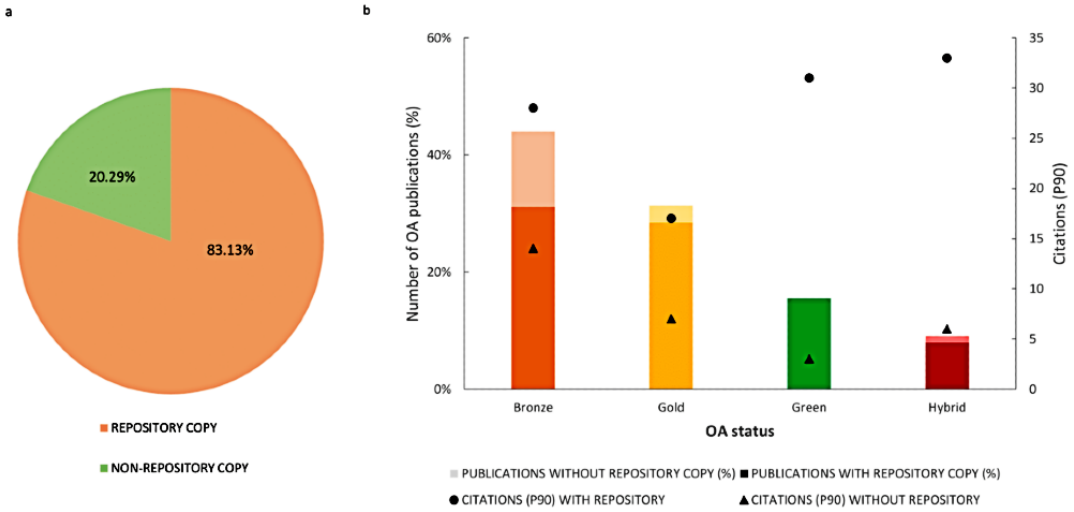


Figure 4 shows the effect of having a repository copy of OA SARS-CoV-2–related papers on citations. As shown in Figure 4a, 83.1% of the OA papers had at least one copy in a repository (70.7% of Bronze OA; 90.7% of Gold OA; 99.9% of Green OA, although one paper was categorized as Green without a repository copy; and 88% of Hybrid OA publications). Among these papers, 37.4% (n=27,990) were categorized as Bronze OA, 34.2% (n=25,583) as Gold OA, 18.7% (n=13,992) as Green OA, and 9.6% (n=7207) as Hybrid OA. More concretely, in every OA typology, the P90 was higher in the group of publications with a repository copy than in the group of those without such a copy: 28 versus 14 for Bronze papers, 17 versus 7 for Gold papers, 31 versus 3 for Green papers, and 33 versus 6 for papers published in Hybrid journals (Figure 4b).

Figure 4. Effect of having a repository copy of open access (OA) SARS-CoV-2–related papers hosted in PubMed (January 1, 2020, to March 1, 2021) on the citations (based on the 90th percentile [P90]). (a) Percentage of OA papers with and without a repository copy. (b) Top 10% of papers with and without a repository copy by OA type.

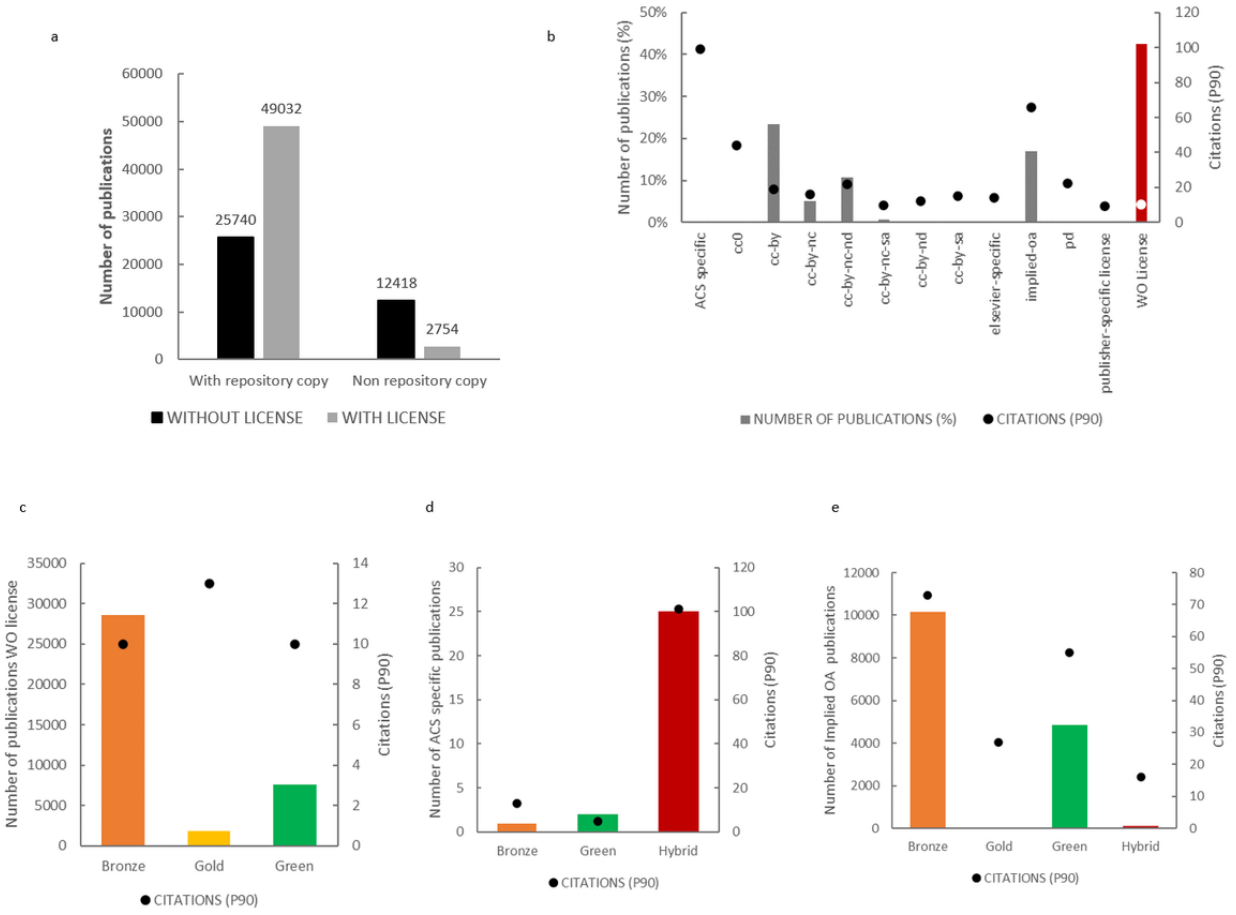


Licenses

We also reviewed the reuse permissions by licenses held by the OA papers: 34.4% (n=25,740) of the papers with a repository copy did not have an explicit license, compared to 81.8% (n=12,418) of those without a repository copy (Figure 5a).

Figure 5b shows that a very relevant number of all OA articles lack a proper license (42.4%), which means licenses allowing free reusability of the paper. The most used licenses are CC-BY (23.3%), followed by Implied-OA (16.9%), CC-BY-NC-ND (10.8%), and CC-BY-NC (5.1%). When the citations of these groups were analyzed, we observed that the highest citation indicator was for papers under ACS-Specific licenses (with 99.1 citations) and Implied-OA licenses (66 citations). Articles without an explicit license showed a poor number of citations (10). Based on these results, these three groups (nonlicensed, ACS-Specific licensed, and Implied-OA licensed) were further studied. For the nonlicensed OA papers, the predominant OA status was Bronze, accounting for 75.1% (n=28,584) of papers with a P90 of 10, followed by Green (20%, P90=10) and Gold (4.9%, P90=13) (Figure 5c). The most cited papers by license type, ACS-Specific licensed papers, were further analyzed. In this case, almost 90% of the papers belonged to the Hybrid OA category with a remarkable P90 value of 101.2 (Figure 5d). Finally, 67.2% of the Implied-OA licensed papers had a Bronze OA status with a P90 value of 73 (Figure 5e).

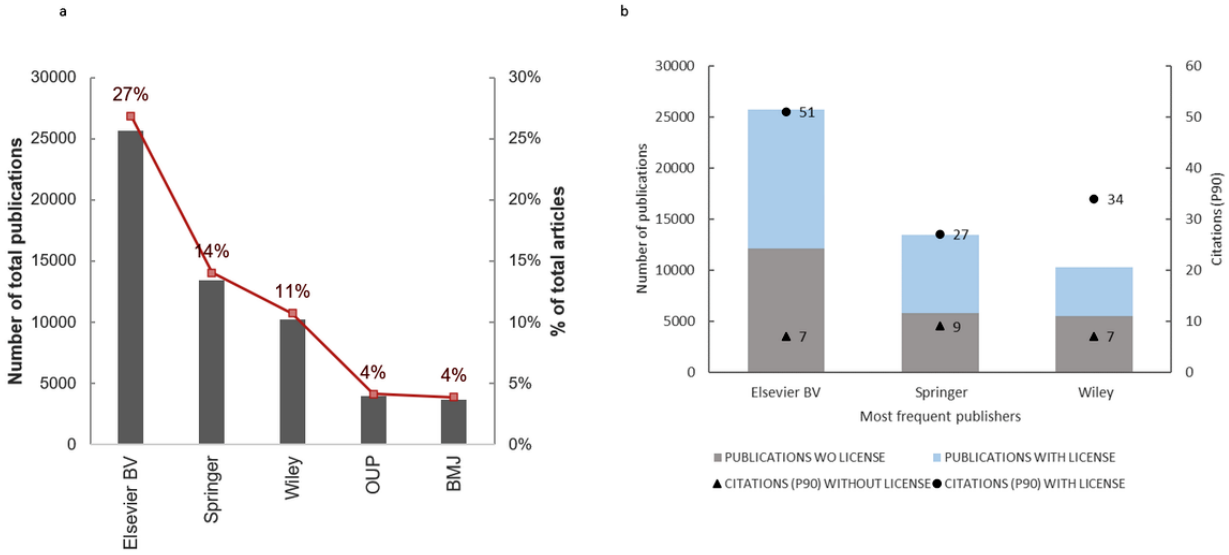
Figure 5. Licensing of open access (OA) SARS-CoV-2–related papers hosted in PubMed (January 1, 2020, to March 1, 2021). (a) Number of papers with and without (WO) a specific licence distributed by OA/non-OA and with/without a repository copy. (b) Distribution of papers based on the licence category. (c-e) P90 and OA status of nonlicensed papers (c), ACS-specific licensed papers (d), and implied OA licensed papers (e). P90: 90th percentile; ACS: American Chemical Society.



Publishers

The most frequent publisher was Elsevier, publishing 26.88% (25,694/95,605) of the included papers, followed by Wiley (13,461/95,605, 14.08%), Springer (10,266/95,605, 10.74%), OUP (3940/95,605, 4.12%), and BMJ (3701/95,605, 3.87%) (Figure 6a). The presence or absence of a certain license for these publishers was studied in greater depth, as well as the citations (P90) of all the publications published by the three top publishers (Figure 6b). The results showed that 47% (n=12,090) of the Elsevier-published papers do not have a license, and the associated number of citations is low (n=7). However, articles from this publisher with a license had a much higher citation P90 of 51. The same pattern was observed for the next two most frequent publishers: 43% of Springer's articles do not have any license and their citation level is low compared to the licensed papers (9 vs 27); 53% of Wiley's papers lack a license and with only 7 citations compared to the 34 citations of the licensed papers.

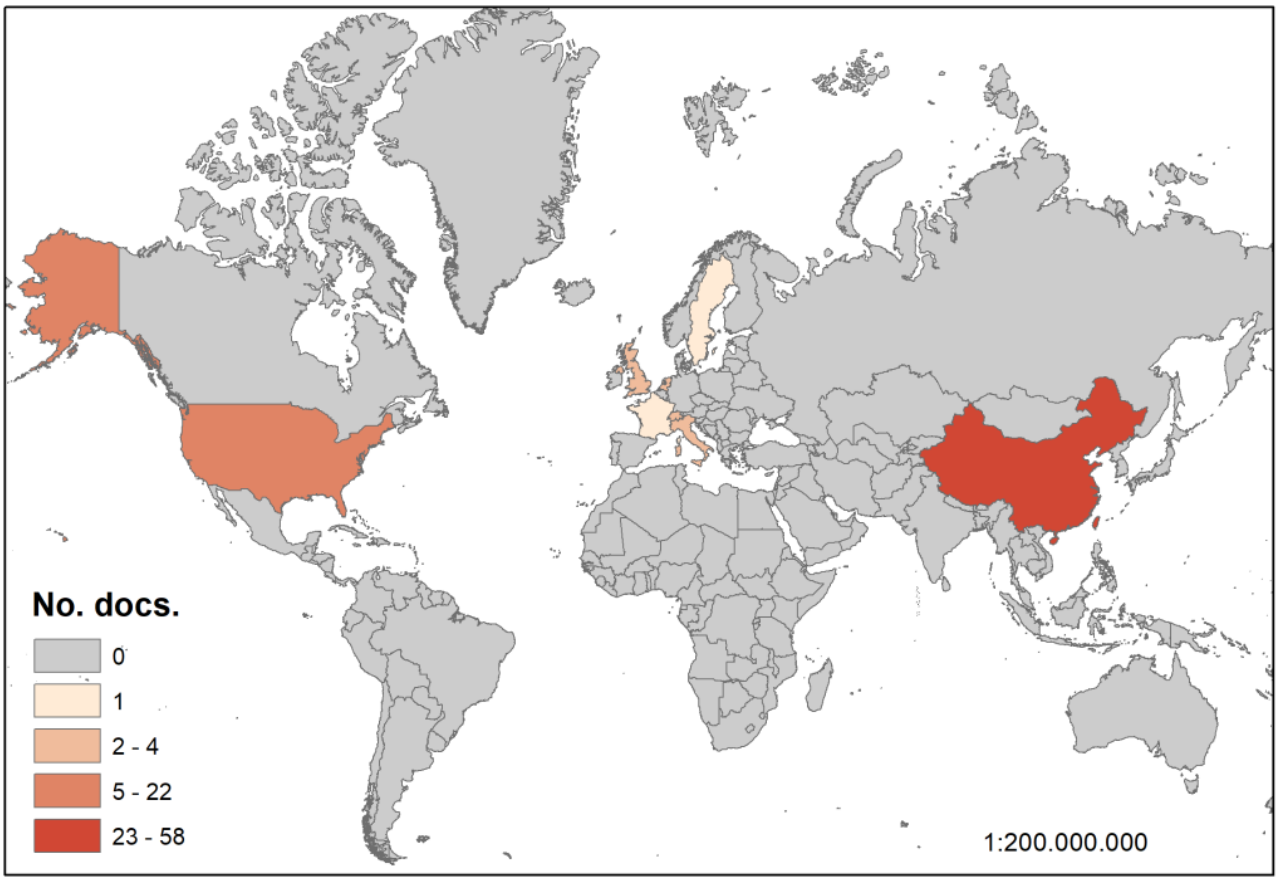
Figure 6. Publishers and journals that published the highest number of COVID-19–related papers hosted by PubMed from January 1, 2020, to March 1, 2021. (a) Number and percentage of total publications distributed by the most frequent publishers. (b) Citation (P90) and presence/absence of a proper licence of all the papers published in the three main publishers. BMJ: British Medical Journal; OUP: Oxford University Press; P90: 90th percentile; WO: without.



Highly Cited Papers by Country

For papers with more than 1000 citations (105 highly cited papers), we determined the country of the corresponding author. China was the country with the most cited papers, including 58 articles with more than 1000 citations (Figure 7). The mean citation value of these 58 papers was 3932, with the highest being 16,164 citations. The two countries with the most highly cited papers were the United States and the United Kingdom, having 22 and 11 papers with more than 1000 citations, respectively. After these three, other countries presented a significantly lower (less than 5) number of highly cited papers (eg, Germany, 4; Italy, the Netherlands, and Switzerland, 2; and France, Singapore, Sweden, and Taiwan, 1).

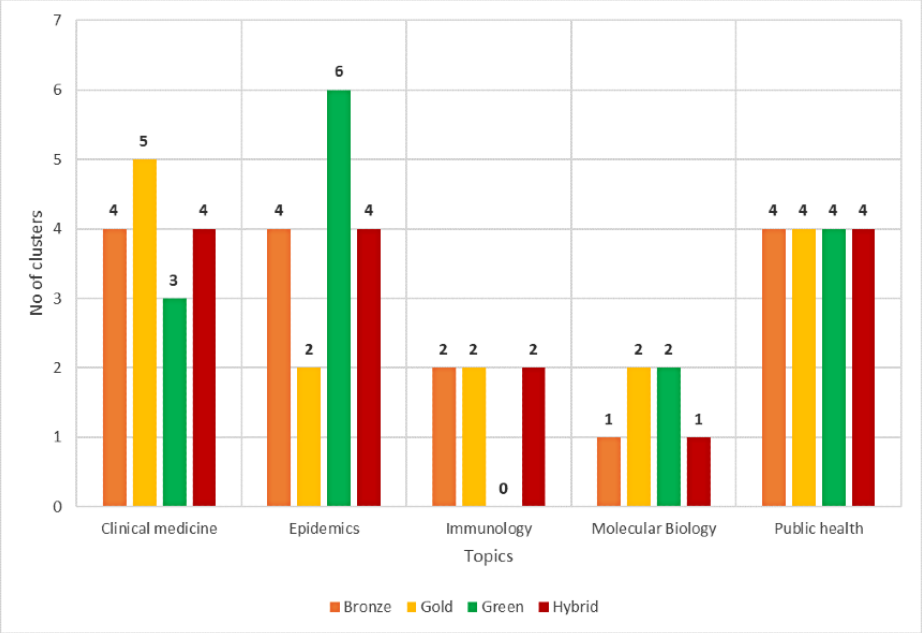
Figure 7. Map of highly cited papers by country of authorship (corresponding author). Image created using ArcGIS [36].



Identifying and Monitoring Topic Evolution

A topic modeling technique based on title and abstracts was used to analyze the biomedical content of each publication together with their distribution during the period studied. Figure 8 shows the number of times that each topic was mentioned by thematic cluster and OA category. Topics such as Public Health, Epidemics (ie, monitoring of COVID-19 within countries), and Clinical Medicine (ie, patients, analysis, therapy) were the most frequently addressed, suggesting that the prevention and control of COVID-19 are the most concerning issues at all stages (see Multimedia Appendix 1). By contrast, Immunology (ie, trials and vaccinations) and Molecular Biology (ie, proteins, antibodies) for the purpose of detection and prevention do not exhibit as much interest. Moreover, some topics show a marked preference for specific OA categories, such as Clinical Medicine in Gold OA and Epidemics in Green OA.

Figure 8. Distribution of the number of COVID-19–related topics by open access type.



Among the Bronze OA publications, as represented in Figure 9, cluster 7 (health care and services) stood out from March 2020. Cluster 3, terms associated with the lockdown and cases (epidemics), was common in January 2020 but decreased over the course of the pandemic. Another prominent cluster was cluster 5, represented by symptoms (eg, respiratory syndrome), which was more common from February 2020 and this popularity was maintained throughout the study period. Similarly, cluster 1, related to general research on COVID-19 (surveys, interviews, etc), gained popularity from April 2020. With a different pattern, cluster 11 (drugs, protein, virus) was relatively common in January 2020 but decreased over the period of study. By contrast, there were some topics with less presence, including clusters 2 and 6, represented by clinical medicine (eg, pregnant women); cluster 4, represented by immunology; and clusters 13 and 14, represented by epidemics (eg, tests and prediction models).

Figure 10 shows the evolution of the topics of Gold OA publications. Cluster 5, related to strategies adopted by countries, stood out throughout the period analyzed. Another relevant topic was the number of cases in China (especially during February 2020) (cluster 9) and clinical symptoms (infection, respiratory syndrome) (cluster 14) during the first months of the pandemic (January-March 2020). Cluster 1 and cluster 8, representing clinical medicine (eg, proteins) and public health (eg, mental health effects of the pandemic), respectively, showed a modest increase during the later months of the study.

Green OA publications are shown in Figure 11. Topics reflected in cluster 6, associated with respiratory symptoms, were very common in January and February 2020. Cluster 5 (treatments for COVID-19, such as hydroxychloroquine) was strong in February 2020. Other evolutions of interest included patients and hospitalization (cluster 10), which gained relevance over time (notably November-December 2021), whereas treatment (cluster 12; eg, drugs, proteins, and antivirals) started being relevant from March to July 2020 and then interest subsequently decreased. Effects (cluster 2; eg, dental, sleep quality) or symptoms and global measures adopted to prevent the virus (cluster 13; eg, lockdown, social distancing) exhibited relatively less interest.

Figure 12 shows the cluster intensity based on the number of Hybrid OA publications over the study period. Clusters 0, 2, and 5 were the most highly studied topics at the beginning of the period analyzed, corresponding to Public Health and Epidemics. As an example, cluster 2 starts with a burst in January 2020 due to the effects of COVID-19 on psychological and mental health (eg, depression, anxiety, psychological effect) of the population. Notably, clusters 3, 6, and 13, associated with the topics Public Health, Clinical Medicine, and Epidemics, respectively, gained intensity over time. Other clusters showing almost no interest were those related with nursing and care (cluster 8), mortality (cluster 11), and child response (cluster 14).

Figure 9. Topic intensity in the Bronze open access journals (January 1, 2020, to March 1, 2021) (n=38,625).

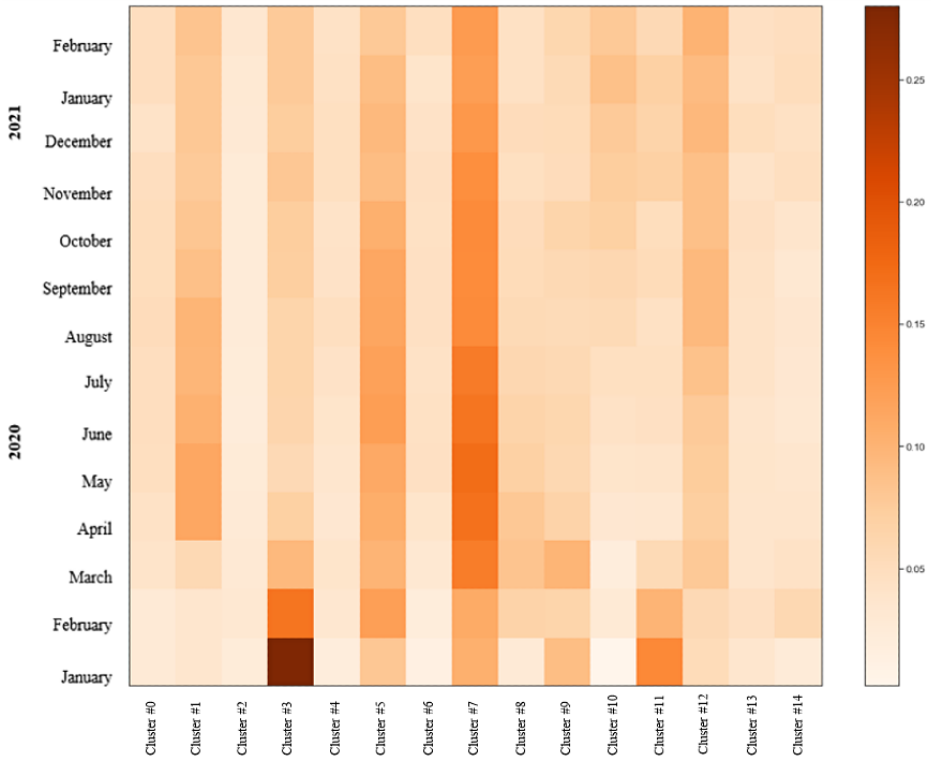


Figure 10. Topic intensity in the Gold open access journals (January 1, 2020, to March 1, 2021) (n=27,786).

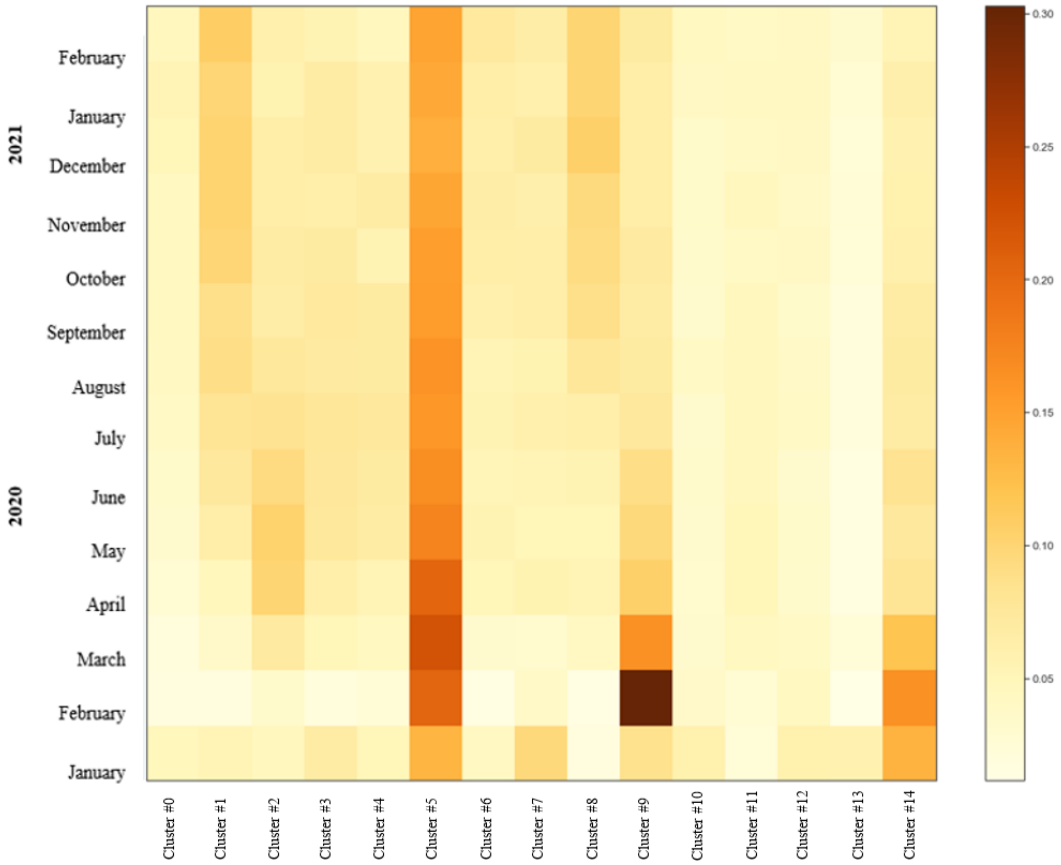


Figure 11. Topic intensity in Green open access journals (January 1, 2020, to March 1, 2021) (n=13,396).

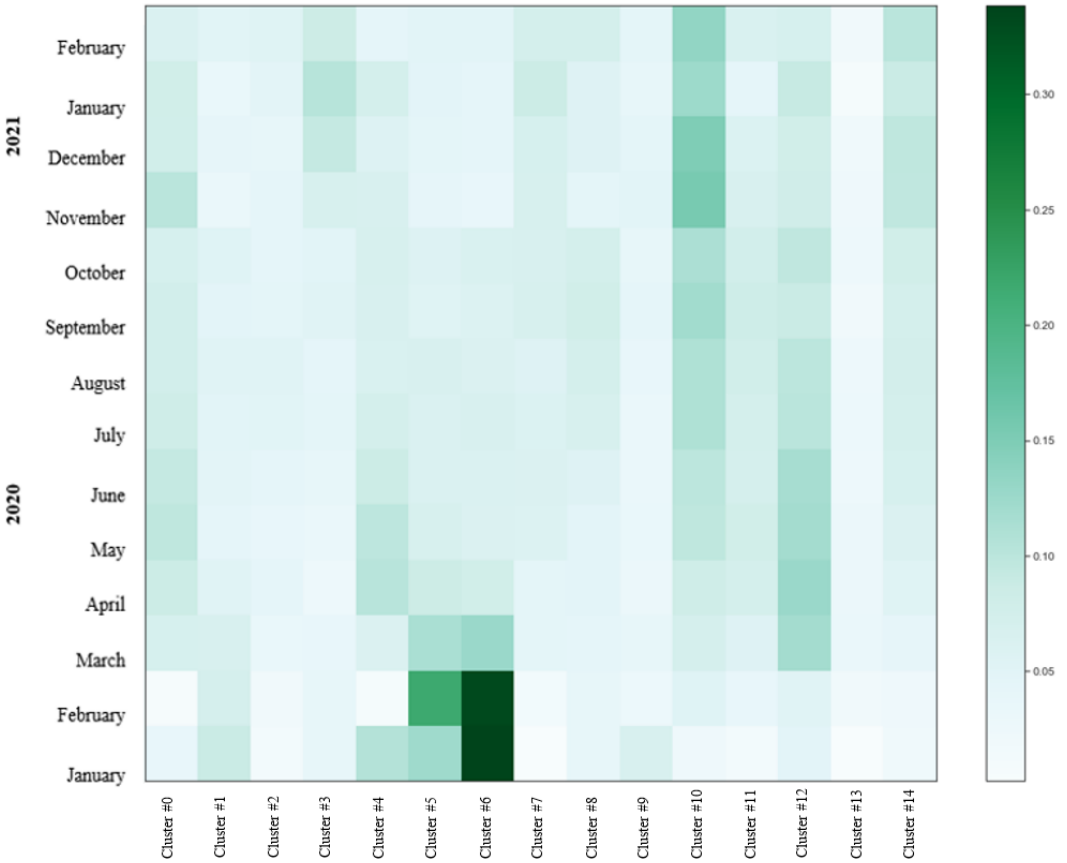
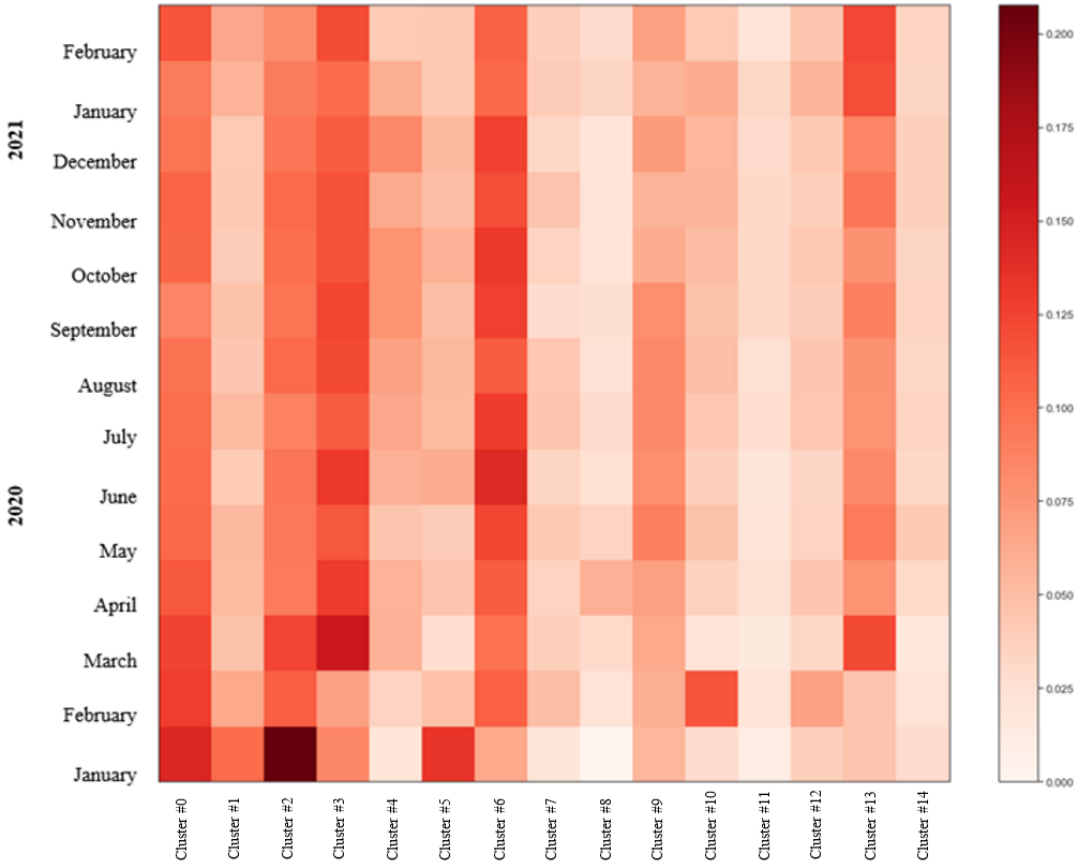


Figure 12. Topic intensity in the Hybrid journals (January 1, 2020, to March 1, 2021) (n=7937).



Discussion

Based on the large increase in the number of publications during the pandemic [15], the data analyzed in this study (95,605 publications) show that the majority of papers are openly available (94.1%), which is a significantly higher rate than found in other databases (eg, 68% in Dimensions, as pointed out by Torres-Salinas et al [5]). Bronze OA was the most common category, which means that paid journals are providing free access for these publications. The same pattern is also supported by previous studies in different databases such as WoS, Scopus, and Dimensions [5,15,37,38]. Analysis of the evolution of the publications and OA types over time showed that, although an increasing tendency is observed in all OA types, Green OA articles decreased in favor of Gold OA journals during the pandemic, in line with the findings of Nane et al [11].

These results highlighted that the OA impact (measured by the P90) is higher in papers with a repository copy; however, 42% of those OA papers do not have a license, which might be correlated with less visibility and could affect the reuse of the findings. Although the most used licenses are CC-BY, Implied-OA, and CC-BY-NC-ND, ACS-Specific and Implied-OA licenses are associated with a higher number of citations. In this regard, if the knowledge and discoveries are not properly shared and transmitted, the struggle against disease is slowed, with more pronounced fatal effects.

The topic modeling analysis showed that the majority of publications in PubMed focus on Public Health, Epidemics, and Clinical Medicine, whereas Immunology and Molecular Biology are the least addressed topics (complementing the findings of Colavizza et al [16] and Wang and Hong [23]). However, topics such as Public Health and Clinical Medicine play a pivotal role (supporting Wang and Hong [23]), providing new insights to those offered by Colavizza et al [16] on the variation on topics in this specific database.

COVID-19 research topics are continuously evolving along with evolution of their publication trends. Overall, prevention and control are the most prevalent topics (in line with Wang and Hong [23]), while prediction (eg, models to forecast) or treatment (eg, drug treatment), or the effects on specific populations (eg, child response, pregnant women) are the least researched topics. The topic intensity over the months of this study presented different behaviors by OA category. Hybrid and Green OA publications are more focused on the patients and their effects, whereas the strategy adopted by different countries is more frequently published in Gold OA journals, and Healthcare and Services topics are largely published in Bronze OA journals. Although the research focus at the beginning of the pandemic was largely concentrated on disease symptoms or treatments to control the spread of the virus (published in Green, Hybrid, and Gold journals), tests or samples (Hybrid), or the number of cases (Gold)—and these topics prevail continuously, such as the public health system in Hybrid journals or strategies from countries in Gold journals—more recently, the focus has been on the cases by country (Hybrid), patients and hospitalization (Green), or proteins (Gold), among others.

The main conclusions of this study can be summarized as follows. First, the number of COVID-19–related articles in PubMed 1 year following the first global lockdown is 17-times higher than that at the initial stage of the pandemic. This provides new insights into the study of Torres-Salinas et al [5], which estimated a total of 1000 documents per week in PubMed at the beginning of the pandemic.

Second, to effectively confront the global pandemic, we need to make research, and its outcomes, more open. This is an opportunity to show how the scholarly communication system can benefit the public. Although a high number of publications are freely available, not all of them are open and reusable. As clearly demonstrated in this study, more effort on public licensing is needed; 42% of the OA papers related to COVID-19 do not have a license, and this is associated with less visibility, especially for Bronze OA publications.

Third, articles with a higher number of citations include those published under journal-imposed licenses that specify that access to these papers is temporary, allowing reuse and analysis for a limited time, or even allowing reading access for a limited time only.

Fourth, as measured by the number of citations, OA categories (specially Hybrid and Green) seem to be associated with a higher impact than closed journals. Even greater impacts are observed with repository copies (especially those with ACS-Specific licenses and Implied-OA licenses).

Fifth, only approximately 100 papers received more than 1000 citations. Papers written in English, from corresponding authors located in developed countries (United States, China, and the United Kingdom) dominate the highly cited papers.

Sixth, Hybrid and Green OA publications are more focused on patients and their effects, whereas the strategy adopted by countries is more prevalent in papers that have chosen the Gold OA route. Health care and services are the most common topics in the papers published in Bronze OA journals.

Finally, prevention and control were the most prevalent topics in the publications analyzed (coronavirus outbreaks/epidemiology and public health). However, research in some topics is still insufficient (eg, effects on some populations such as children or pregnant women), requiring more global research collaborations.

Overall, monitoring and measuring OA and topic evolution will help researchers and scientific policymakers understand the status of COVID-19 research. This information may be useful as a reference guide, to stimulate new ideas and directions of research, and to help in the fight against this pandemic.

Acknowledgments

The authors would like to acknowledge the thoughtful review and feedback from Charles McCathieNevile to this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clusters of COVID-19–related publications according to open access type based on topic modeling.
[DOCX File , 24 KB - jmir_v24i9e40011_app1.docx]

References

1. Khachfe H, Chahrour M, Sammouri J, Salhab H, Makki B, Fares M. An epidemiological study on COVID-19: a rapidly spreading disease. Cureus 2020 Mar 18;12(3):e7313 [FREE Full text] [doi: 10.7759/cureus.7313] [Medline: 32313754]
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
3. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf [accessed 2022-08-23]
4. Weekly epidemiological update on COVID-19 - 30 March 2021. World Health Organization. 2021 Mar 28. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---31-march-2021 [accessed 2022-08-23]
5. Torres-Salinas D, Robinson-Garcia N, Castillo-Valdivieso P. Open access and altmetrics in the pandemic age: forescast analysis on COVID-19 literature. bioRxiv. 2020 Apr 26. URL: https://www.biorxiv.org/content/10.1101/2020.04.23.057307v1 [accessed 2022-08-23]
6. Kupferschmidt K. Preprints bring 'firehose' of outbreak data. Science 2020 Feb 28;367(6481):963-964. [doi: 10.1126/science.367.6481.963] [Medline: 32108094]
7. VOSviewer Visualizing scientific landscapes. URL: https://www.vosviewer.com/ [accessed 2022-09-12]
8. Xie B, He D, Mercer T, Wang Y, Wu D, Fleischmann KR, et al. Global health crises are also information crises: a call to action. J Assoc Inf Sci Technol 2020 Dec;71(12):1419-1423 [FREE Full text] [doi: 10.1002/asi.24357] [Medline: 32427189]
9. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598. [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
10. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Wellcome. 2020 Jan 31. URL: https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak [accessed 2022-08-23]
11. Nane T, Robinson-Garcia N, van SF, Torres-Salinas D. COVID-19 and the scientific publishing system: Growth, open access and scientific fields. SocArXiv Papers. 2021 Dec 15. URL: https://osf.io/preprints/socarxiv/ntrpx/ [accessed 2022-08-23]
12. Horbach SPJM. No time for that now! Qualitative changes in manuscript peer review during the Covid-19 pandemic. Res Eval 2021;30(3):231-239 [FREE Full text] [doi: 10.1093/reseval/rvaa037]
13. Vervoort D, Ma X, Shrime MG. Money down the drain: predatory publishing in the COVID-19 era. Can J Public Health 2020 Oct 04;111(5):665-666 [FREE Full text] [doi: 10.17269/s41997-020-00411-5] [Medline: 32886305]
14. Lariviere V, Shu F, Sugimoto C. The Coronavirus (COVID-19) outbreak highlights serious deficiencies in scholarly communication. LSE Impact Blog. 2020 Mar 05. URL: https://blogs.lse.ac.uk/impactofsocialsciences/2020/03/05/the-coronavirus-covid-19-outbreak-highlights-serious-deficiencies-in-scholarly-communication/ [accessed 2022-08-23]
15. Arrizabalaga O, Otaegui D, Vergara I, Arrizabalaga J, Méndez E. Open access of COVID-19-related publications in the first quarter of 2020: a preliminary study based in PubMed. F1000Res 2020;9:649 [FREE Full text] [doi: 10.12688/f1000research.24136.2] [Medline: 32850121]
16. Colavizza G, Costas R, Traag VA, van Eck NJ, van Leeuwen T, Waltman L. A scientometric overview of CORD-19. PLoS One 2021 Jan 7;16(1):e0244839 [FREE Full text] [doi: 10.1371/journal.pone.0244839] [Medline: 33411846]
17. Kousha K, Thelwall M. COVID-19 publications: database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. Quant Sci Stud 2020 Aug;1(3):1068-1091. [doi: 10.1162/qss_a_00066]
18. Colavizza G. COVID-19 research in Wikipedia. Quant Sci Stud 2020 Dec;1(4):1349-1380. [doi: 10.1162/qss_a_00080]
19. Thelwall M. Coronavirus research before 2020 is more relevant than ever, especially when interpreted for COVID-19. Quant Sci Stud 2020 Dec;1(4):1381-1395. [doi: 10.1162/qss_a_00083]
20. Cunningham E, Smyth B, Greene D. Collaboration in the time of COVID: a scientometric analysis of multidisciplinary SARS-CoV-2 research. Humanit Soc Sci Commun 2021 Oct 19;8(1):240. [doi: 10.1057/s41599-021-00922-7]

21. Damaševičius R, Zailskaitė-Jakštė L. Impact of COVID-19 pandemic on researcher collaboration in business and economics areas on national level: a scientometric analysis. J Doc 2022 Jun 07:online ahead of print. [doi: 10.1108/jd-02-2022-0030]
22. Viglione G. Are women publishing less during the pandemic? Here's what the data say. Nature 2020 May 20;581(7809):365-366. [doi: 10.1038/d41586-020-01294-9] [Medline: 32433639]
23. Wang J, Hong N. The COVID-19 research landscape: Measuring topics and collaborations using scientific literature. Medicine 2020 Oct 23;99(43):e22849. [doi: 10.1097/MD.0000000000022849] [Medline: 33120818]
24. Homolak J, Kodvanj I, Virag D. Preliminary analysis of COVID-19 academic information patterns: a call for open science in the times of closed borders. Scientometrics 2020 Jun 25;124(3):2687-2701 [FREE Full text] [doi: 10.1007/s11192-020-03587-2] [Medline: 32836524]
25. Soltani P, Patini R. Retracted COVID-19 articles: a side-effect of the hot race to publication. Scientometrics 2020 Aug 05;125(1):819-822 [FREE Full text] [doi: 10.1007/s11192-020-03661-9] [Medline: 32836531]
26. Waltman L, Pinfield S, Rzayeva N, Henriques SO, Fang Z, Brumberg J, et al. Scholarly communication in times of crisis: the response of the scholarly communication system to the COVID-19 pandemic. Research on Research Institute. 2021. URL: https://rori.figshare.com/articles/report/Scholarly_communication_in_times_of_crisis_The_response_of_the_scholarly_communication_system_to_the_COVID-19_pandemic/17125394 [accessed 2022-08-23]
27. Falagas M, Pitsouni E, Malietzis G, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2007 Sep 20;22(2):338-342. [doi: 10.1096/fj.07-9492lsf]
28. Our World in Data. URL: https://ourworldindata.org/coronavirus#explore-the-global-situation [accessed 2022-08-23]
29. Rodrigo C. Acceso abierto (Open Access) en las publicaciones científicas. Revista Española de Nutrición Comunitaria (Spanish Journal of Community Nutrition) 2010;16(4):203.
30. Hofmann T. Probabilistic latent semantic analysis. 1999 Presented at: 15th Conference on Uncertainty in Artificial Intelligence (UAI); July 30-August 1, 1999; Stockholm, Sweden.
31. Blei D, Ng A, Jordan M. Latent dirichllocation. J Machine Learn Res 2003;3:993-1022.
32. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
33. Feinerer I, Hornik K. tm: Text mining package version 0.7-8. R Project. 2020. URL: https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf [accessed 2022-08-23]
34. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. 2010 Presented at: LREC 2010 Workshop on New Challenges for NLP Frameworks; May 22, 2010; Malta. [doi: 10.13140/2.1.2393.1847]
35. San Torcuato M, Bautista-Puig N, Arrizabalaga O, Mendez E. Tracking openness and topic evolution of COVID-19 publications: a comprehensive analysis (January 2020- March 2021). Zenodo. 2022 Mar 22. URL: https://zenodo.org/record/6376578#.YwTw_OzMJcA [accessed 2022-08-23]
36. ArcGIS Online. URL: https://www.arcgis.com/index.html [accessed 2022-09-12]
37. Chen X, Bharti N, Marsteller MR. Use of bibliometrics data to understand the citation advantages of different open access categories in Covid-19 related studies. Proc Assoc Inf Sci Technol 2021;58(1):410-414 [FREE Full text] [doi: 10.1002/pra2.469] [Medline: 34901398]
38. Belli S, Mugnaini R, Baltà J, Abadal E. Coronavirus mapping in scientific publications: when science advances rapidly and collectively, is access to this knowledge open to society? Scientometrics 2020;124(3):2661-2685 [FREE Full text] [doi: 10.1007/s11192-020-03590-7] [Medline: 32836526]

Abbreviations

ACS: American Chemical Society
APC: article processing charge
BMJ: British Medical Journal
CC: Creative Commons
MeSH: Medical Subject Heading
OA: open access
OIA-PMH: Open Archives Initiative-Protocol for Metadata Harvesting
OUP: Oxford University Press
P90: 90th percentile
WoS: Web of Science

Edited by C Basch; submitted 01.06.22; peer-reviewed by J Willinsky, R Damaševičius; comments to author 24.06.22; revised version received 21.07.22; accepted 28.07.22; published 03.10.22.

Please cite as:
San Torcuato M, Bautista-Puig N, Arrizabalaga O, Méndez E
Tracking Openness and Topic Evolution of COVID-19 Publications January 2020-March 2021: Comprehensive Bibliometric and Topic Modeling Analysis
J Med Internet Res 2022;24(10):e40011
URL: https://www.jmir.org/2022/10/e40011
doi:10.2196/40011
PMID:36190742

Original Paper

Influence of Digital Intervention Messaging on Influenza Vaccination Rates Among Adults With Cardiovascular Disease in the United States: Decentralized Randomized Controlled Trial

Nell J Marshall[1*], MPH, DrPH; Jennifer L Lee[1*], PhD; Jessica Schroeder[1], PhD; Wei-Nchih Lee[1], MPH, MD, PhD; Jermyn See[1], PhD; Mohammad Madjid[2], MS, MD; Mrudula R Munagala[3], MS, MD; John D Piette[4], MS, PhD; Litjen Tan[5], MS, PhD; Orly Vardeny[6], MS, PharmD; Michael Greenberg[7], MPH, MD; Jan Liska[8], MsC; Monica Mercer[7], MD; Sandrine Samson[9], PhD

[1]Evidation Health, Inc, San Mateo, CA, United States
[2]David Geffen School of Medicine, University of California, Los Angeles, CA, United States
[3]Department of Cardiology, University of Miami, Miami, FL, United States
[4]Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor, MI, United States
[5]Immunize.org, St. Paul, MN, United States
[6]Center for Care Delivery and Outcomes Research, Veterans Health Administration, Minneapolis, MN, United States
[7]Sanofi, Swiftwater, PA, United States
[8]Sanofi, Gentilly, France
[9]Sanofi, Lyon, France
[*]these authors contributed equally

Corresponding Author:
Nell J Marshall, MPH, DrPH
Evidation Health, Inc
63 Bovet Rd #146
San Mateo, CA, 94402
United States
Phone: 1 (415) 515 1985
Email: nmarshall@evidation.com

Abstract

Background: Seasonal influenza affects 5% to 15% of Americans annually, resulting in preventable deaths and substantial economic impact. Influenza infection is particularly dangerous for people with cardiovascular disease, who therefore represent a priority group for vaccination campaigns.

Objective: We aimed to assess the effects of digital intervention messaging on self-reported rates of seasonal influenza vaccination.

Methods: This was a randomized, controlled, single-blind, and decentralized trial conducted at individual locations throughout the United States over the 2020-2021 influenza season. Adults with self-reported cardiovascular disease who were members of the Achievement mobile platform were randomized to receive or not receive a series of 6 patient-centered digital intervention messages promoting influenza vaccination. The primary end point was the between-group difference in self-reported vaccination rates at 6 months after randomization. Secondary outcomes included the levels of engagement with the messages and the relationship between vaccination rates and engagement with the messages. Subgroup analyses examined variation in intervention effects by race. Controlling for randomization group, we examined the impact of other predictors of vaccination status, including cardiovascular condition type, vaccine drivers or barriers, and vaccine knowledge.

Results: Of the 49,138 randomized participants, responses on the primary end point were available for 11,237 (22.87%; 5575 in the intervention group and 5662 in the control group) participants. The vaccination rate was significantly higher in the intervention group (3418/5575, 61.31%) than the control group (3355/5662, 59.25%; relative risk 1.03, 95% CI 1.004-1.066; *P*=.03). Participants who were older, more educated, and White or Asian were more likely to report being vaccinated. The intervention was effective among White participants (*P*=.004) but not among people of color (*P*=.42). The vaccination rate was 13 percentage points higher among participants who completed all 6 intervention messages versus none, and at least 2 completed messages appeared to be

needed for effectiveness. Participants who reported a diagnosis of COVID-19 were more likely to be vaccinated for influenza regardless of treatment assignment.

Conclusions: This personalized, evidence-based digital intervention was effective in increasing vaccination rates in this population of high-risk people with cardiovascular disease.

Trial Registration: ClinicalTrials.gov NCT04584645; https://clinicaltrials.gov/ct2/show/NCT04584645

(J Med Internet Res 2022;24(10):e38710) doi:10.2196/38710

KEYWORDS

influenza; randomized trial; public health; cardiovascular disease; immunization; vaccination; digital messaging; digital intervention; mobile health; mHealth

Introduction

About 5% to 15% of the US population contracts influenza annually [1], resulting in more than 20,000 deaths [1] and substantial economic impact [2]. For people with cardiovascular disease (CVD), influenza can be particularly dangerous. In one study, the risk of myocardial infarction was 6 times higher within a week of influenza infection [3]. A study of more than 80,000 US adults hospitalized with influenza over 8 seasons found that 1 in every 8 patients developed sudden, serious cardiac complications and that having underlying cardiac disease was significantly associated with experiencing an acute cardiac event with influenza [4]. For these reasons, the Centers for Disease Control and Prevention (CDC) consider persons with CVD to be at high risk for influenza complications and therefore a priority group for vaccination [5].

Vaccination remains the most effective primary prevention method against influenza, with age-adjusted effectiveness rates of up to 68% over the past 5 years [6]. The CDC reported a 51.4% vaccination rate for 2019-2020 for persons aged 18-64 years who have high-risk conditions such as CVD [7], far below the 70% national vaccination rate goal [8]. Given the increased burden of influenza for people with CVD, even small improvements in vaccination rates could substantially reduce the number of patients having major adverse cardiac events [2,9-11].

Novel, scalable, cost-optimal, and effective solutions are needed to address barriers to influenza vaccination among people with CVD, such as complacency, time and cost constraints, and a lack of confidence [12]. Observational [13] and randomized controlled trials [14,15] have shown the effectiveness of digital messaging to increase vaccine uptake in general adult populations. In a randomized trial of digital messaging in persons with diabetes, a population also at increased risk of influenza-related complications [16], the vaccination rate was 3.1% higher in the intervention group than the control group. Alternatively stated, 33 people would need to receive the intervention for 1 additional person to become vaccinated.

The primary objective of this study was to examine the efficacy of a digital intervention designed to increase self-reported influenza vaccination rates in individuals with CVD.

Methods

Study Design

This 8-month, pragmatic randomized controlled trial was conducted remotely in the United States. Participants were blinded to study participation status to minimize observation bias, although all participants agreed that their survey responses and behavioral data would be used for research purposes before completing each survey (see below).

All participants were members of the free Achievement mobile health and research platform (Evidation Health, Inc), which includes more than 4 million individuals spanning all 50 states and 90% of zip codes [17]. The platform provides personalized insights and tools to motivate and empower people to take evidence-supported actions to manage their health. Members can connect activity trackers and fitness and health apps to the platform and share self-reported health information. Achievement does not have the ability to access clinical or claims data; it relies solely on member-generated data.

Ethics Approval

The trial protocol was approved by Solutions IRB, Yarnell, Arizona (Registration: IORG0007116; Federalwide Assurance: IRB00008523), and registered at ClinicalTrials.gov (NCT04584645). Since the digital intervention messages were consistent with publicly available information on influenza vaccination, we obtained a waiver of informed consent from this Institutional Review Board on the basis that participants would face only minimal risk from the study. Participants were informed about how their survey responses and behavioral data would be used through a Data Usage and Permissions Agreement.

Digital Intervention Design

The 6 digital intervention messages were developed using a 3-part approach [18], building on a previous study [16] and the Theory of Planned Behavior [19]. Message designs were refined throughout the development process using Rapid Iterative Testing and Evaluation–inspired methods [20]. See Multimedia Appendix 1 [16,18-21] for details of the development process and the content of the intervention messages (Table S1 and Figures S1-7 in Multimedia Appendix 1), which were delivered via the Achievement platform.

Participants

Eligible participants were those aged ≥18 years, living in the United States, with any of the following self-reported conditions on the Achievement platform (eg, through past surveys): atrial fibrillation; abnormal or irregular heart rhythm or other arrhythmic heart disease; cardiac arrest or myocardial infarction; coronary artery disease treated with medication, stenting, percutaneous intervention, or bypass surgery; congestive heart failure; or stroke or cerebrovascular accident.

Recruitment, Screening, and Enrolling

Members who met the inclusion criteria were identified for study inclusion ("participants"). Participants took no action to enroll and were not informed about their participation status. We used block randomization by cardiovascular condition to randomize participants into either the intervention group, which received the digital intervention messages, or the control group, which received none of the messages.

Randomization and Blinding

Evidation Health, Inc generated the random allocations, enrolled participants, and randomized them using block randomization (arrhythmia vs nonarrhythmia) into the intervention or control group before offering the opportunity to complete any study activities.

Study Procedures

Participants were asked to complete the web-based surveys at baseline, 3 months (after 4 digital messages had been sent in the intervention group), and 6 months (after 2 more messages had been sent in the intervention group). Reminder messages were used to motivate survey completion.

Primary and Other Outcomes

Participants self-reported their vaccination status (yes or no) via the app at baseline, 3 months, and 6 months. Participants also reported the estimated date of vaccination, if any, on the 3- and 6-month surveys.

To assess engagement with the intervention messages, we examined platform-generated data indicating that the person had completed a given message and created a summary measure indicating the number of messages completed.

Each survey measured the drivers and barriers to vaccination as well as vaccine knowledge. The vaccine drivers and barriers of interest included the number of visits to a primary care provider in the 3 months before randomization (none, 1-2, or 3 or more), number of visits to a cardiology specialist in the prior 3 months (none, 1-2, or 3 or more), number of hospitalizations in the prior 3 months (none, 1, or 2 or more), whether a health care provider had offered influenza vaccination (yes, no, or unsure), and whether a health care provider had informed the individual that they were in a "high-risk group" (yes, no, or unsure).

Vaccine knowledge factors were based on responses to the survey question "What sources of information do you use to learn about the flu vaccine?" with possible responses of health care professionals, family member or peers, social media including blog posts, mobile apps, or conventional news media (eg, television and newspapers).

Sample Size Calculation

The sample size was determined a priori for a 2-arm interventional statistical superiority study design with self-reported vaccination rates as the primary outcome [22]. Large studies on the impact of messaging and telephone reminders to improve influenza vaccination rates show a range of effect sizes from 2.5% to 3.5% [23,24]. A total of 8000 participants were needed to detect a 3% difference in vaccination rates with a type I error rate of 0.05 and power of 0.80. Since a participation drop-off of 67% was observed for digital interventions aimed at increasing influenza vaccination in people with diabetes [16], we conservatively estimated an engagement rate of about 16%. The targeted enrollment list therefore included approximately 49,000 individuals to yield the analysis population of 8000 participants.

Statistical Analysis

We first compared the unadjusted proportions of participants reporting vaccination at follow-up between the intervention and control groups. In predefined subgroup analyses, we examined variations in intervention effects between White and non-White participants. Process analyses included differences in self-reported vaccination rates within the intervention group by the number of intervention messages completed and intervention participants' levels of engagement with each message. Controlling for randomization group, we examined other predictors of vaccination status, including cardiovascular condition type, vaccine drivers or barriers, and vaccine knowledge.

An exploratory objective was to describe the impact of the COVID-19 pandemic on influenza vaccination behavior. Another exploratory objective—self-reported complications from influenza, overall and by vaccination status—was not analyzed because the surveys did not ask about influenza complications. Information on safety and adverse events was not collected, given the minimal-risk nature of the intervention and study.

Variables were compared at the 5% significance level using 2-sided tests or 2-sided 95% CI unless otherwise specified. Comparison of means used 2-sided Student *t* test for normal distributions or a Mann-Whitney *U* test for nonnormal distributions. Comparisons of frequencies used chi-square tests. For the logistic regression model, the *P* values, odds ratios (ORs), and 95% CIs associated with each of the β parameter estimates were reported. To describe the relative importance of each predictor variable, we calculated their Shapley Additive Explanation values [25]. Kaplan-Meier curves were constructed for time to influenza vaccination, using the participant-estimated dates of influenza vaccination from the 3- and 6-month surveys.

Results

Participants

Between July and September 2020, we generated a list of 49,138 candidate participants (Figure 1). Of these, 24,570 were

randomized to receive digital intervention messages and 24,568 were randomized to the control group. On September 21, 2020, the first baseline and demographic surveys were sent to these 49,138 participants, and 10,402 (21.17%) completed the baseline survey. In all, 11,237 participants (22.87%) completed the midstudy or final survey by April 11, 2021, yielding groups of 5575 intervention and 5662 control participants who reported vaccination status at either 3 or 6 months after randomization.

Of the 11,237 participants, the average age was 45 (SD 13) years, 81.18% (n=9122) were White, 78.01% (n=8766) were female, and 86.21% (n=9687) had health insurance (Table 1). More than half (n=6891, 61.32%) had a college degree, and a third (n=3770, 33.55%) had a household income of at least US $75,000. The most commonly reported cardiovascular condition was arrhythmia (intervention: 2251/5575, 40.38%; control: 2331/5662, 41.17%). Baseline characteristics did not differ substantially between groups. Despite previous self-reports of CVD from all participants, almost a third (intervention: 1798/5575, 32.35%; control: 1844/5662, 32.57%) in both groups reported not having any of the listed conditions in the baseline survey. Study participants represented all 50 states and the District of Columbia (Figure S8 in Multimedia Appendix 1).

Figure 1. Disposition of Study Participants.

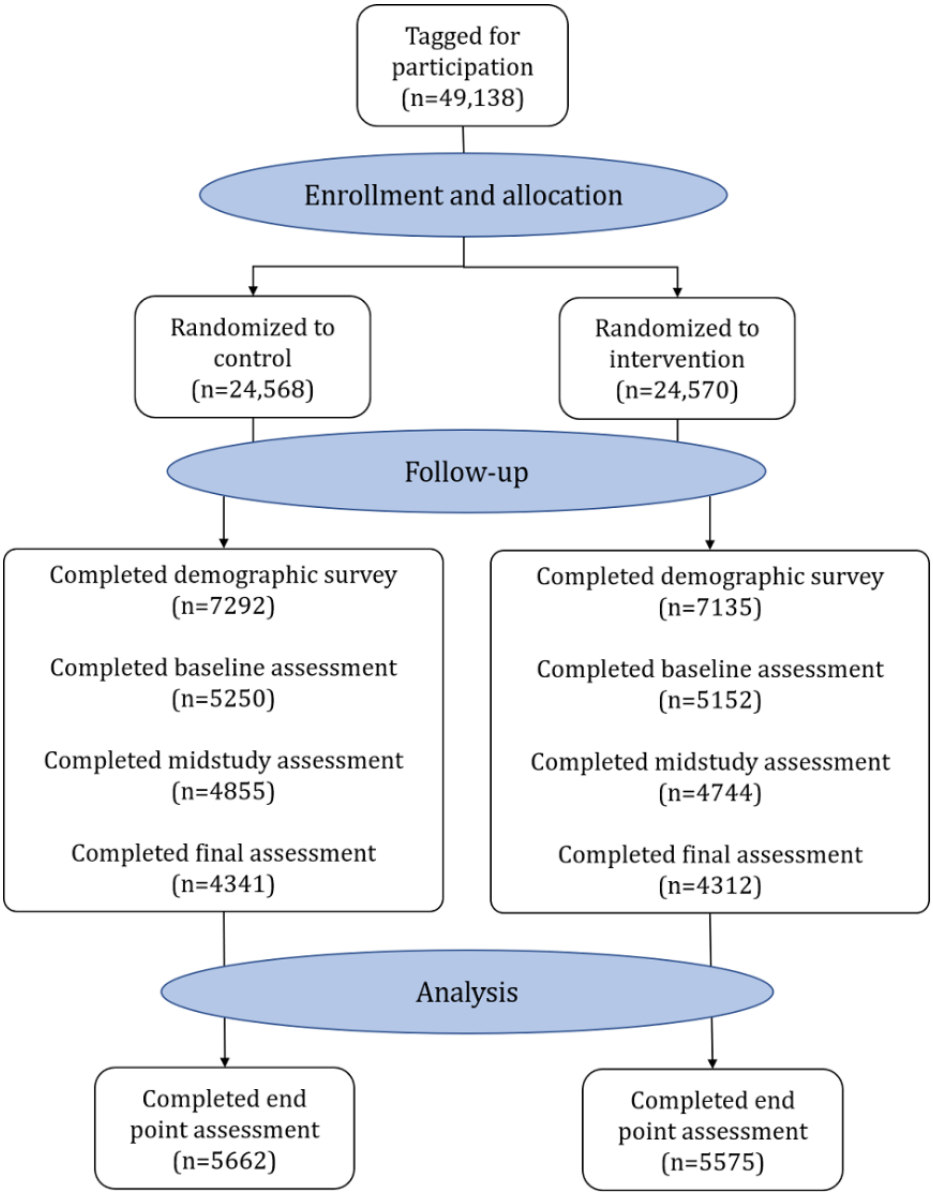


Table 1. Baseline demographic and clinical characteristics of the participants.

Characteristic	Intervention (N=5575)	Control (N=5662)
Age (years; intervention: n=5530; control: n=5607), mean (SD)	45.0 (13.5)	44.9 (13.3)
Sex, n (%)[a]		
Female	4339 (77.83)	4427 (78.19)
Male	1071 (19.21)	1114 (19.68)
Other	169 (3.03)	135 (2.38)
Race/ethnicity, n (%)[a]		
American Indian or Alaska Native	155 (2.78)	143 (2.53)
Asian	250 (4.48)	227 (4)
Black or African American	388 (6.96)	354 (6.25)
Hispanic, Latino, or Spanish	319 (5.72)	375 (6.62)
Middle Eastern or North African	46 (0.83)	50 (0.88)
Native Hawaiian or Other Pacific Islander	45 (0.81)	49 (0.87)
White	4510 (80.9)	4612 (81.46)
Other	60 (1.08)	64 (1.13)
Prefer not to answer	187 (3.35)	179 (3.16)
Had health insurance, n (%)	4815 (86.37)	4872 (86.05)
Had a college degree, n (%)	3381 (60.65)	3510 (61.99)
Household income ≥US $75,000, n (%)	1861 (33.38)	1909 (33.72)
Cardiovascular condition type, n (%)[a]		
Arrhythmia	2251 (40.37)	2331 (41.17)
Atrial fibrillation	488 (8.75)	496 (8.76)
Cardiac arrest	112 (2)	99 (1.75)
Myocardial infarction	385 (6.91)	412 (7.28)
Heart failure	332 (5.96)	293 (5.17)
Coronary artery disease	366 (6.57)	356 (6.29)
Stroke or cerebrovascular accident	433 (7.77)	436 (7.7)
Other cardiovascular diseases	539 (9.67)	545 (9.63)
None of the above diagnoses[b]	1798 (32.25)	1844 (32.57)

[a]Participants could choose more than 1 option, and percentages may add up to >100%.

[b]Despite previous self-reports of cardiovascular disease from all participants, some reported not having any of the included conditions at baseline. Please see the Limitations section for more details.

Primary Outcome

By the end of the study period, 3418 (61.31%) of the 5575 participants in the intervention arm had reported obtaining influenza vaccination compared to 3355 (59.25%) of the 5662 participants in the control arm (absolute difference: 2.06%; relative risk 1.03, 95% CI 1.004-1.066; *P*=.03). Based on this difference, 48.3 persons would have to receive the digital intervention messages for 1 additional person to become vaccinated.

Secondary Outcomes

In logistic regression modeling, overall predictors of vaccination status included White or Asian race and being older or a college graduate (Figure 2 and Figure S9 in Multimedia Appendix 1). Being in the intervention group was associated with a significantly increased likelihood of getting the influenza vaccine (OR 1.099, 95% CI 1.012-1.192; *P*=.02). Participants who had cardiac arrest (OR 3.477, 95% CI 1.85-6.54; *P*<.001), atrial fibrillation (OR 1.332, 95% CI 1.068-1.66; *P*=.01), or coronary disease (OR 1.411, 95% CI 1.055-1.885; *P*=.02) were also more likely to report vaccination (>65%) than participants with other conditions. Digital interventions appeared to be more effective in encouraging vaccinations among White participants (intervention: 2837/4510, 62.9% vs control: 2763/4612, 59.91%; *P*=.004) than among non-White participants (intervention: 581/1065, 54.55% vs control: 593/1050, 56.48%; *P*=.42; Figure S10 in Multimedia Appendix 1).

Kaplan-Meier analysis of the time to vaccination showed that at least 2 digital intervention messages, completed 2 weeks apart, were necessary for a difference to begin to emerge (Figure 3). In the intervention group (N=5575), the most completed messages were the knowledge quiz (n=4248, 76.2%), cost article (n=4276, 76.7%), and CDC article (n=4315, 77.4%; Figure S11 in Multimedia Appendix 1). In all, 44.81% (n=2498) of the intervention group completed all 6 messages, and 7.7% (n=429) completed none of them; the reported vaccination rate for the former group was about 13 percentage points higher than that for the latter group (1626/2498, 65.09% vs 223/429, 51.98%).

Figure 2. Predictors of self-reported influenza vaccination.

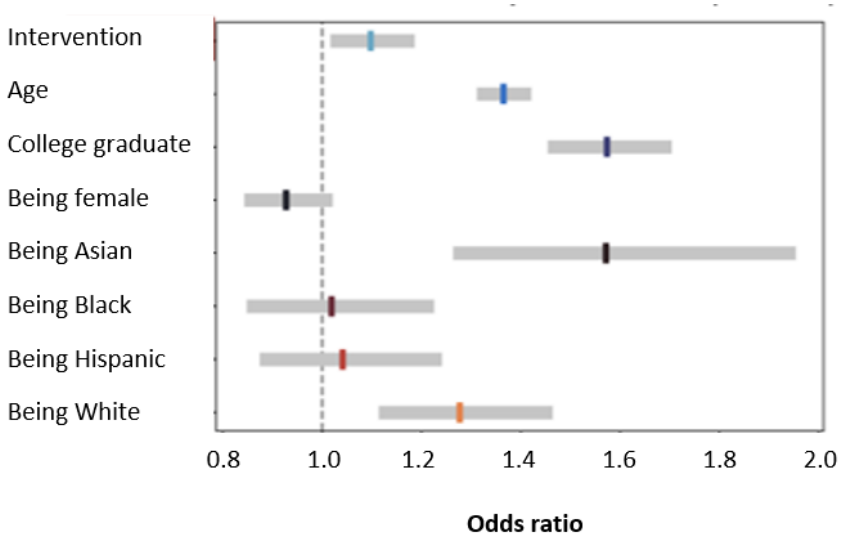
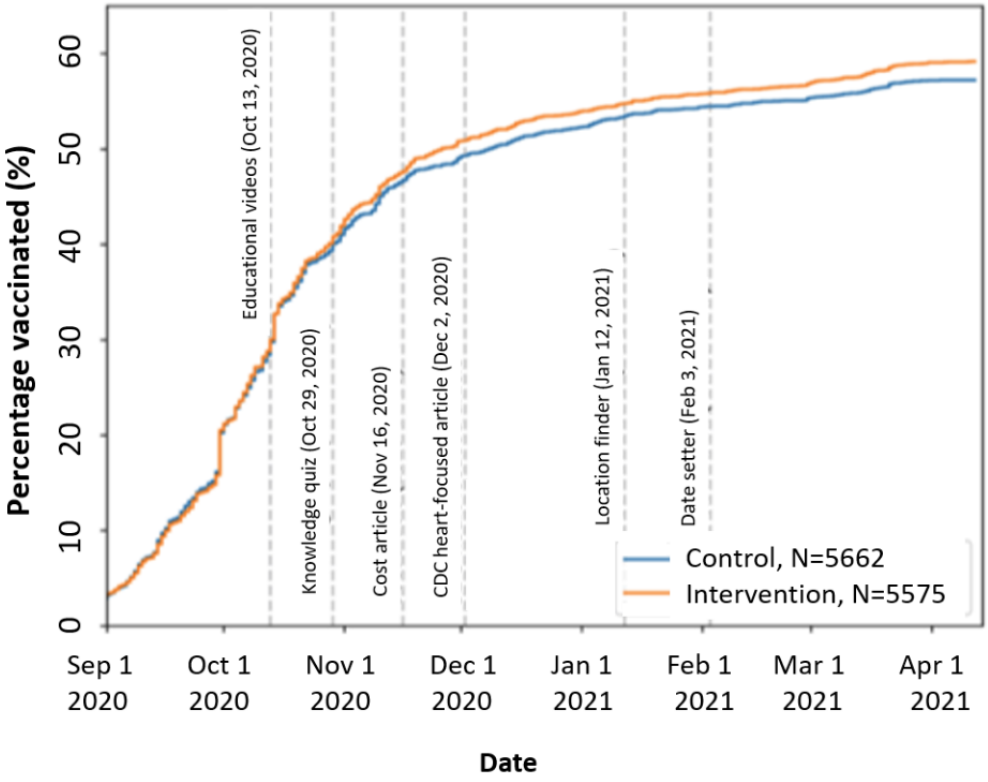


Figure 3. Self-reported vaccination rates over time. Dashed vertical gray lines indicate the timings of the 6 digital intervention messages. CDC: Centers for Disease Control and Prevention.



After controlling for age, sex, race, and education over the entire study population, participants who saw a health care provider, were offered an influenza vaccine, or got their vaccine information from a health care provider were more likely to report getting the vaccine (Figure S12 in Multimedia Appendix 1). Those who were told by a health care provider that they were part of a high-risk group were also more likely to report vaccination (OR 2.369, 95% CI 2.171-2.586; *P*<.001).

Participants who reported a diagnosis of COVID-19 were 40% more likely to report influenza vaccination than those who did not, regardless of intervention assignment (Figure S13 in Multimedia Appendix 1). Of the 7457 participants who reported getting the influenza vaccine, 4252 (57.02%) stated that the COVID-19 pandemic did not influence their decision to vaccinate, and 4026 (66.33%) of the 6070 participants who did not report influenza vaccination said their decision was not influenced by the COVID-19 pandemic.

Discussion

Principal Findings

In this cohort of 11,237 adults with CVD, digital intervention was associated with a significantly higher rate of self-reported influenza vaccination at the end of the study period than control participants. Based on epidemiologic estimates, roughly 26 million Americans have CVD [26], and an increase in vaccination rates of 2.06% as shown in this trial would mean another 535,600 persons with CVD being immunized. This increase would likely translate to substantial reductions in morbidity, mortality, and costs to the health care system, as well as potential improvements in the quality of life if applied at scale.

Our findings add to a growing body of evidence that interventions delivered via digital communication channels can be effective in improving vaccination rates among high-risk patients. Previous randomized studies have generally shown significantly improved influenza vaccination rates with email prompts, app-based messages, SMS text messaging, and web-based interventions in general adult populations [14,15,23,27-31], high-risk patients (some of whom had heart disease) [16,32], and pregnant women [33-35], but a few have not [36-38]. To our knowledge, this is the first randomized study to show promising results with a digital intervention specifically designed for and delivered to a population with CVD.

The patient-centered digital interventions, developed with evidence-based behavioral theory of vaccine behavior [19], were generally viewed as informative, trustworthy, and engaging. In all, about 45% of the intervention group completed all intervention messages, compared to 27% in our previous study [16] and the industry standard of 22% [39], indicating very strong engagement. These messages also produced results at least as good as other recent app-based digital influenza vaccination interventions in general Canadian [13] and US adult populations [14,15].

Older age, more education, and White or Asian race were significant predictors of vaccination in this study. The apparent lack of effect in other participants of color might reflect small sample sizes or heterogeneity among non-White participants. Further assessment of racial and ethnic differences in responses to nontailored digital interventions is needed. This study does reinforce the importance of engagement with the health care system, as participants who saw a health care provider, were told they were in a high-risk group, or were offered the influenza vaccine were more likely to report vaccination.

Vaccination against influenza is a cost-effective method for reducing some of the risk associated with CVD [40,41]. Combined with the cost-effectiveness of digital intervention design and deployment relative to other prevention strategies, the messaging presented here appears to be suitable as a population-health management strategy in the context of limited budgets for health systems, insurers, and public agencies. In addition, half of the 380,000 people [1] hospitalized annually with influenza in the United States have heart disease [5]. Scaling this digital intervention to the larger population of people with CVD could help reduce hospitalizations and emergency department and clinic visits, along with days of productivity lost, particularly in already digitally connected populations.

The strengths of this study include its decentralized, pragmatic nature, which can provide high-quality evidence of effectiveness in real-world settings. Other strengths imclude its large sample size, nationwide scope, and variety of data collected, including patient-generated health data. The study also reflects real-world data on vaccination rates among persons with variable risk levels from influenza infection conferred by different cardiovascular diagnoses. The design of the study may inspire the design of future vaccination campaigns to assess the drivers of vaccination and their public health impact and investigate vaccination behavior in other patient populations.

Limitations and Future Work

Participants reported their CVD diagnoses at different times via different survey sources (eg, historical surveys vs current self-reports). This method resulted in discrepancies from using different data sources in health outcomes, potentially due to question formatting and the time period for recall: almost a third of participants reported having none of the candidate CVD conditions despite previous self-reports of such disease. Future studies could forgo blinding in favor of supplementing self-reports with additional sources of information (eg, health claims and medical records). Participants were blinded to participation, reflecting real-world engagement with health messages outside of a known research-related setting. The potential influence of unknowingly participating in research is unclear. Only about 23% of the sample reported on the primary end point. The generalizability of findings to nonresponders is therefore limited. We also have no knowledge about why participants did not respond.

This trial was conducted during the COVID-19 pandemic. Due to the pandemic, participants may have had increased awareness of viral diseases and vaccines generally through other sources (eg, governmental sources, television, and social media), possibly limiting the generalizability of our findings, although most participants in both groups stated that the pandemic did not affect their decision about influenza vaccination.

Participants in the intervention group were compensated in the form of points, which could be redeemed for cash. However, given that the total possible monetary compensation was only US $1.52 regardless of vaccination status, it likely did not influence the motivation to vaccinate enough to impact the outcome.

All participants were existing members of the Achievement platform, reflecting a population already engaged with digital technology. The baseline (control) vaccination rate (59.25%) was also about 8% higher than the 51% CDC average for individuals with comorbidities [7]. Thus, it might have been more difficult to see an incremental uplift compared to populations with less technology use or a lower baseline vaccination rate.

Most of the population was female, non-Hispanic, and White. The effects of the intervention in other demographic groups are less certain, although the sample size was sufficient for models adjusting for age, education, sex, and race to confirm that the intervention effect remained significant. Barriers to health equity in accessing digital health interventions and methodologies remain significant [42]. This study should serve as a foundation for future evaluation and tailoring to reach individuals from diverse backgrounds more effectively, as Brewer and colleagues [43] have shown that people from diverse racial and ethnic backgrounds engage with digital health information via the web and digital health research at a high rate.

Although several evidence-based sources and techniques were leveraged in the development of the intervention messages, their exact mechanisms of action are unknown. The act of prompting, rather than the content, might result in similar improvement. Future studies examining which components or messages would be the most beneficial could help optimize future interventions while minimizing burden.

Conclusions

A digital intervention using health condition–relevant information and widely available public health information can be an effective way to increase influenza vaccination rates in persons with CVD. These results may have broader public health implications as an easily scalable intervention to increase vaccination behavior. Future studies should examine the effectiveness and cost-effectiveness of such digital campaigns in diverse populations with other chronic conditions and for other types of vaccination, such as COVID-19 vaccines.

Acknowledgments

This study was funded by Sanofi. This sponsor was involved in the intervention development, study design, data interpretation, and writing of the report and was kept informed during data collection and data analysis. Sanofi produces one of the influenza vaccines available in the United States, but participants were not instructed to get a specific type of influenza vaccine. The senior (SS) and corresponding (NJM) authors had full access to all of the data in the study and had final responsibility for the decision to submit for publication. The authors thank Patricia A. French, BS, of Left Lane Communications for assistance in the drafting and editing of the manuscript.

Data Availability

Qualified researchers may request access to the aggregate results and related study documents including the study report, study protocol with any amendments, blank case report form, statistical analysis plan, and data set specifications. Further details on Sanofi's data sharing criteria, eligible studies, and process for requesting access can be found at https://www.vivli.org/.

Authors' Contributions

NJM, JLL, J Schroeder, WNL, J See, M Madjid, MRM, LT, OV, MG, JL, M Mercer, and SS conceived the study and contributed to its design. NJM, JLL, J Schroeder, WNL, J See, JDP, and LT were responsible for the acquisition, analysis, or interpretation of data. NJM, JLL, J Schroeder, WNL, J See, and JD contributed to the drafting of the manuscript. J See conducted the statistical analyses. All authors critically revised the manuscript for important intellectual content and approved the final manuscript.

Conflicts of Interest

NM, J Schroeder, WNL, and J See are current or former employees of Evidation Health, Inc, the developer of the Achievement health and research platform, and may hold stock options in Evidation Health, Inc. JLL is a former employee of Evidation Health, Inc and a current employee of Lyra Health, Inc; receives income from Lyra Health, Inc; and has been granted equity in Lyra Health, Inc. M Madjid has received consulting fees from Sanofi and Seqirus. MRM has received consultation fees, honoraria, or both from Sanofi, CareDx, Alnylam, and Akcea. OV has received consulting fees from Sanofi. MG, M Mercer, SS, and JL are employees of Sanofi, the study sponsor. JL and SS are also shareholders of Sanofi. No other authors declare a conflict of interest.

Multimedia Appendix 1
Supplemental information.
[PDF File (Adobe PDF File), 723 KB - jmir_v24i10e38710_app1.pdf]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) eHealth checklist.
[PDF File (Adobe PDF File), 88 KB - jmir_v24i10e38710_app2.pdf]

References

1. Estimated flu-related illnesses, medical visits, hospitalizations, and deaths in the United States — 2019-2020 flu season. Centers for Disease Control and Prevention. 2021 Oct 01. URL: https://www.cdc.gov/flu/about/burden/2019-2020.html [accessed 2022-04-06]
2. Putri WCWS, Muscatello DJ, Stockwell MS, Newall AT. Economic burden of seasonal influenza in the United States. Vaccine 2018 Jun 22;36(27):3960-3966. [doi: 10.1016/j.vaccine.2018.05.057] [Medline: 29801998]
3. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. N Engl J Med 2018 Jan 25;378(4):345-353. [doi: 10.1056/NEJMoa1702090] [Medline: 29365305]
4. Chow EJ, Rolfes MA, O'Halloran A, Anderson EJ, Bennett NM, Billing L, et al. Acute cardiovascular events associated with influenza in hospitalized adults: a cross-sectional study. Ann Intern Med 2020 Oct 20;173(8):605-613 [FREE Full text] [doi: 10.7326/M20-1509] [Medline: 32833488]
5. Flu and people with heart disease or history of stroke. Centers for Disease Control and Prevention. 2021 Aug 26. URL: https://www.cdc.gov/flu/highrisk/heartdisease.htm [accessed 2022-04-06]
6. US flu vaccine effectiveness data for 2019-2020. Centers for Disease Control and Prevention. 2020 Dec 22. URL: https://www.cdc.gov/flu/vaccines-work/2019-2020.html [accessed 2022-04-06]
7. Flu vaccination coverage, by age group and high risk status, adults, United States, 2019-20 season. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/flu/excel/fluvaxview/2020/general_population/OnlineReport_2019_20_AdditionalTable_FluVxByAge_Adults.xlsx [accessed 2022-04-06]
8. Increase the proportion of people who get the flu vaccine every year—IID-09. Healthy People 2030. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/vaccination/increase-proportion-people-who-get-flu-vaccine-every-year-iid-09 [accessed 2022-04-06]
9. MacIntyre CR, Mahimbo A, Moa AM, Barnes M. Influenza vaccine as a coronary intervention for prevention of myocardial infarction. Heart 2016 Dec 15;102(24):1953-1956 [FREE Full text] [doi: 10.1136/heartjnl-2016-309983] [Medline: 27686519]
10. Yedlapati SH, Khan SU, Talluri S, Lone AN, Khan MZ, Khan MS, et al. Effects of influenza vaccine on mortality and cardiovascular outcomes in patients with cardiovascular disease: a systematic review and meta-analysis. J Am Heart Assoc 2021 Mar 16;10(6):e019636 [FREE Full text] [doi: 10.1161/JAHA.120.019636] [Medline: 33719496]
11. Behrouzi B, Udell JA. Moving the needle on atherosclerotic cardiovascular disease and heart failure with influenza vaccination. Curr Atheroscler Rep 2021 Oct 21;23(12):78 [FREE Full text] [doi: 10.1007/s11883-021-00973-w] [Medline: 34671861]
12. Schmid P, Rauber D, Betsch C, Lidolt G, Denker M. Barriers of influenza vaccination intention and behavior - a systematic review of influenza vaccine hesitancy, 2005-2016. PLoS One 2017;12(1):e0170550 [FREE Full text] [doi: 10.1371/journal.pone.0170550] [Medline: 28125629]
13. Dale LP, White L, Mitchell M, Faulkner G. Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake. Vaccine 2019 Jul 26;37(32):4594-4600. [doi: 10.1016/j.vaccine.2018.04.018] [Medline: 29699784]
14. Lee W, Stück D, Konty K, Rivers C, Brown CR, Zbikowski SM, et al. Large-scale influenza vaccination promotion on a mobile app platform: a randomized controlled trial. Vaccine 2020 Apr 16;38(18):3508-3514. [doi: 10.1016/j.vaccine.2019.11.053] [Medline: 31787410]
15. Milkman KL, Gandhi L, Patel MS, Graci HN, Gromet DM, Ho H, et al. A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. Proc Natl Acad Sci U S A 2022 Feb 08;119(6):e2115126119 [FREE Full text] [doi: 10.1073/pnas.2115126119] [Medline: 35105809]
16. Lee JL, Foschini L, Kumar S, Juusola J, Liska J, Mercer M, et al. Digital intervention increases influenza vaccination rates for people with diabetes in a decentralized randomized trial. NPJ Digit Med 2021 Sep 17;4(1):138 [FREE Full text] [doi: 10.1038/s41746-021-00508-2] [Medline: 34535755]
17. Evidation Health, Inc. URL: https://evidation.com/ [accessed 2021-04-06]
18. Schroeder J, Madjid M, Munagala M, Piette J, Tan LJ, Vardeny O, et al. Developing a patient-centered digital intervention for increasing influenza vaccination rates in people with cardiovascular disease using a remote mixed-methods approach [poster]. 2021 Dec Presented at: Eighth European Scientific Working group on Influenza (ESWI) Influenza Conference; December 6-9, 2021; Valencia, Spain.
19. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
20. Medlock MC, Wixon D, Terrano M, Romero RL, Fulton B. Using the RITE method to improve products: a definition and a case study. 2002 Presented at: UPA 2002: 11th Annual Usability Professionals' Association Conference; July 8-12, 2002; Orlando, FL URL: https://www.jpattonassociates.com/wp-content/uploads/2015/04/rite_method.pdf
21. Frank L, Basch E, Selby JV, Patient-Centered Outcomes Research Institute. The PCORI perspective on patient-centered outcomes research. JAMA 2014 Oct 15;312(15):1513-1514. [doi: 10.1001/jama.2014.11100] [Medline: 25167382]
22. Zhong B. How to calculate sample size in randomized controlled trial? J Thorac Dis 2009 Dec;1(1):51-54 [FREE Full text] [Medline: 22263004]

23. Nehme EK, Delphia M, Cha EM, Thomas M, Lakey D. Promoting influenza vaccination among an ACA health plan subscriber population: a randomized trial. Am J Health Promot 2019 Jul;33(6):916-920. [doi: 10.1177/0890117118823157] [Medline: 30630342]
24. Hurley LP, Beaty B, Lockhart S, Gurfinkel D, Breslin K, Dickinson M, et al. RCT of centralized vaccine reminder/recall for adults. Am J Prev Med 2018 Aug;55(2):231-239. [doi: 10.1016/j.amepre.2018.04.022] [Medline: 29910118]
25. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv 2017 May 22:1-10. [doi: 10.48550/arXiv.1705.07874]
26. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 update: a report from the American Heart Association. Circulation 2021 Feb 23;143(8):e254-e743 [FREE Full text] [doi: 10.1161/CIR.0000000000000950] [Medline: 33501848]
27. Baskin E. Increasing influenza vaccination rates via low cost messaging interventions. PLoS One 2018;13(2):e0192594 [FREE Full text] [doi: 10.1371/journal.pone.0192594] [Medline: 29444147]
28. Wijesundara JG, Ito Fukunaga M, Ogarek J, Barton B, Fisher L, Preusse P, et al. Electronic health record portal messages and interactive voice response calls to improve rates of early season influenza vaccination: randomized controlled trial. J Med Internet Res 2020 Sep 25;22(9):e16373 [FREE Full text] [doi: 10.2196/16373] [Medline: 32975529]
29. Cutrona SL, Golden JG, Goff SL, Ogarek J, Barton B, Fisher L, et al. Improving rates of outpatient influenza vaccination through EHR portal messages and interactive automated calls: a randomized controlled trial. J Gen Intern Med 2018 May;33(5):659-667 [FREE Full text] [doi: 10.1007/s11606-017-4266-9] [Medline: 29383550]
30. Szilagyi PG, Albertin C, Casillas A, Valderrama R, Duru OK, Ong MK, et al. Effect of patient portal reminders sent by a health care system on influenza vaccination rates: a randomized clinical trial. JAMA Intern Med 2020 Jul 01;180(7):962-970 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1602] [Medline: 32421168]
31. Cheung YA. Improving influenza vaccination rates in nurses through electronic text message reminders: a randomized controlled trial [thesis]. The University of Hong Kong. 2016. URL: https://hub.hku.hk/handle/10722/241308 [accessed 2022-08-16]
32. Regan AK, Bloomfield L, Peters I, Effler PV. Randomized controlled trial of text message reminders for increasing influenza vaccination. Ann Fam Med 2017 Nov;15(6):507-514 [FREE Full text] [doi: 10.1370/afm.2120] [Medline: 29133488]
33. Bushar JA, Kendrick JS, Ding H, Black CL, Greby SM. Text4baby influenza messaging and influenza vaccination among pregnant women. Am J Prev Med 2017 Dec;53(6):845-853 [FREE Full text] [doi: 10.1016/j.amepre.2017.06.021] [Medline: 28867143]
34. O'Leary ST, Narwaney KJ, Wagner NM, Kraus CR, Omer SB, Glanz JM. Efficacy of a web-based intervention to increase uptake of maternal vaccines: an RCT. Am J Prev Med 2019 Oct;57(4):e125-e133. [doi: 10.1016/j.amepre.2019.05.018] [Medline: 31471001]
35. Stockwell MS, Westhoff C, Kharbanda EO, Vargas CY, Camargo S, Vawdrey DK, et al. Influenza vaccine text message reminders for urban, low-income pregnant women: a randomized controlled trial. Am J Public Health 2014 Feb;104 Suppl 1:e7-12. [doi: 10.2105/AJPH.2013.301620] [Medline: 24354839]
36. Herrett E, Williamson E, van Staa T, Ranopa M, Free C, Chadborn T, et al. Text messaging reminders for influenza vaccine in primary care: a cluster randomised controlled trial (TXT4FLUJAB). BMJ Open 2016 Feb 19;6(2):e010069 [FREE Full text] [doi: 10.1136/bmjopen-2015-010069] [Medline: 26895984]
37. Yudin MH, Mistry N, De Souza LR, Besel K, Patel V, Blanco Mejia S, et al. Text messages for influenza vaccination among pregnant women: a randomized controlled trial. Vaccine 2017 Feb 01;35(5):842-848. [doi: 10.1016/j.vaccine.2016.12.002] [Medline: 28062124]
38. Moniz MH, Hasley S, Meyn LA, Beigi RH. Improving influenza vaccination rates in pregnancy through text messaging: a randomized controlled trial. Obstet Gynecol 2013 Apr;121(4):734-740. [doi: 10.1097/AOG.0b013e31828642b1] [Medline: 23635672]
39. Email marketing benchmarks and statistics by industry. Mailchimp. 2019 Oct. URL: https://mailchimp.com/resources/email-marketing-benchmarks/ [accessed 2022-04-06]
40. Udell JA, Zawi R, Bhatt DL, Keshtkar-Jahromi M, Gaughran F, Phrommintikul A, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis. JAMA 2013 Oct 23;310(16):1711-1720. [doi: 10.1001/jama.2013.279206] [Medline: 24150467]
41. Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. Cochrane Database Syst Rev 2015 May 05(5):CD005050 [FREE Full text] [doi: 10.1002/14651858.CD005050.pub3] [Medline: 25940444]
42. Wood BR, Young JD, Abdel-Massih RC, McCurdy L, Vento TJ, Dhanireddy S, et al. Advancing digital health equity: a policy paper of the Infectious Diseases Society of America and the HIV Medicine Association. Clin Infect Dis 2021 Mar 15;72(6):913-919 [FREE Full text] [doi: 10.1093/cid/ciaa1525] [Medline: 33033829]
43. 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]

Abbreviations

CDC: Centers for Disease Control and Prevention
CVD: cardiovascular disease
OR: odds ratio

Edited by A Mavragani; submitted 13.04.22; peer-reviewed by R Rahimi, H Veldandi; comments to author 01.06.22; revised version received 21.06.22; accepted 31.07.22; published 07.10.22.

Please cite as:
Marshall NJ, Lee JL, Schroeder J, Lee WN, See J, Madjid M, Munagala MR, Piette JD, Tan L, Vardeny O, Greenberg M, Liska J, Mercer M, Samson S
Influence of Digital Intervention Messaging on Influenza Vaccination Rates Among Adults With Cardiovascular Disease in the United States: Decentralized Randomized Controlled Trial
J Med Internet Res 2022;24(10):e38710
URL: https://www.jmir.org/2022/10/e38710
doi:10.2196/38710
PMID:36206046

Original Paper

A Smartphone App to Promote Patient Activation and Support Shared Decision-making in People With a Diagnosis of Schizophrenia in Outpatient Treatment Settings (Momentum Trial): Randomized Controlled Assessor-Blinded Trial

Tobias Vitger[1,2], PhD; Carsten Hjorthøj[3,4], PhD; Stephen F Austin[5], PhD; Lone Petersen[1], PhD; Esben Sandvik Tønder[6], MA; Merete Nordentoft[2,3], MD, PhD; Lisa Korsbek[7], PhD

[1]Competence Center for Rehabilitation and Recovery, Mental Health Center Ballerup, Mental Health Services in the Capital Region of Denmark, Ballerup, Denmark

[2]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[3]Copenhagen Research Center for Mental Health – CORE, Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark

[4]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[5]Psychiatric Research Unit, Psychiatry Region Zealand, Slagelse, Denmark

[6]The Mental Health Services of Zealand, Slagelse, Denmark

[7]The Mental Health Centre Odense, Mental Health Services in the Region of Southern Denmark, Odense, Denmark

Corresponding Author:
Tobias Vitger, PhD
Competence Center for Rehabilitation and Recovery
Mental Health Center Ballerup
Mental Health Services in the Capital Region of Denmark
Maglevænget 19
Ballerup, 2750
Denmark
Phone: 45 51408519
Email: tobias246@hotmail.com

Abstract

Background: Shared decision-making (SDM) is a process aimed at facilitating patient-centered care by ensuring that the patient and provider are actively involved in treatment decisions. In mental health care, SDM has been advocated as a means for the patient to gain or regain control and responsibility over their life and recovery process. To support the process of patient-centered care and SDM, digital tools may have advantages in terms of accessibility, structure, and reminders.

Objective: In this randomized controlled trial, we aimed to investigate the effect of a digital tool to support patient activation and SDM.

Methods: The trial was designed as a randomized, assessor-blinded, 2-armed, parallel-group multicenter trial investigating the use of a digital SDM intervention for 6 months compared with treatment as usual. Participants with a diagnosis of schizophrenia, schizotypal or delusional disorder were recruited from 9 outpatient treatment sites in the Capital Region of Denmark. The primary outcome was the self-reported level of activation at the postintervention time point. The secondary outcomes included self-efficacy, hope, working alliance, satisfaction, preparedness for treatment consultation, symptom severity, and level of functioning. Explorative outcomes on the effect of the intervention at the midintervention time point along with objective data on the use of the digital tool were collected.

Results: In total, 194 participants were included. The intention-to-treat analysis revealed a statistically significant effect favoring the intervention group on patient activation (mean difference 4.39, 95% CI 0.99-7.79; Cohen *d*=0.33; *P*=.01), confidence in communicating with one's provider (mean difference 1.85, 95% CI 0.01-3.69; Cohen *d*=0.24; *P*=.05), and feeling prepared for decision-making (mean difference 5.12, 95% CI 0.16-10.08; Cohen *d*=0.27; *P*=.04). We found no effect of the digital SDM tool on treatment satisfaction, hope, self-efficacy, working alliance, severity of symptoms, level of functioning, use of antipsychotic medicine, and number or length of psychiatric hospital admissions.

Conclusions: This trial showed a significant effect of a digital SDM tool on the subjective level of patient activation, confidence in communicating with one's provider, and feeling prepared for decision-making at the postintervention time point. The effect size was smaller than the 0.42 effect size that we had anticipated and sampled for. The trial contributes to the evidence on how digital tools may support patient-centered care and SDM in mental health care.

Trial Registration: ClinicalTrials.gov NCT03554655; https://clinicaltrials.gov/ct2/show/NCT03554655

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-019-2143-2

(J Med Internet Res 2022;24(10):e40292) doi:10.2196/40292

KEYWORDS

mobile health; mHealth; digital intervention; shared decision-making; patient activation; schizophrenia; schizotypal; early intervention; randomized clinical trial; mobile phone

Introduction

Shared Decision-making

Shared decision-making (SDM) is a collaborative process between ≥2 partners. In a health care setting, SDM is often designated to be between a patient and provider. It is a continuous cycle aimed at facilitating patient-centered care and making joint treatment decisions. In mental health care, SDM has been proposed as a means to contribute to recovery-oriented care by inviting the patient to have more control and be more involved in their treatment decisions [1].

The current evidence on the effectiveness of SDM in mental health care is somewhat inconclusive but appears to be promising. Studies have found that SDM interventions improve self-perceived involvement in decision-making [2], satisfaction [2], therapeutic alliance [2], decision self-efficacy [3], and adherence to pharmacological treatment [4].

Incorporating SDM into daily practice in mental health care has shown to face some of the same barriers as recovery-oriented interventions, such as changing health care professionals' paternalistic approaches, beliefs that SDM is time consuming and inappropriate for patients with severe mental illness, or discrepancies between the patients' needs and values versus the goals and values of the health care provider and the organization [5,6].

To address these barriers, providers are encouraged to consider the patients' decision-making skills, talk to the patient about how they prefer a decision process to be, and incorporate tools to support the SDM process [7]. In addition, activating patients may also support SDM; active patients who seek collaborative care could also activate their provider, resulting in a good foundation for SDM [8]. Much research has been conducted on patient activation with the conceptualization that active patients consider their own role in the treatment to be important, are engaged in managing their own health and care, feel confident when collaborating with their provider, and have the knowledge and skills to manage their condition [9]. The ability to maintain these behaviors even during stressful times is believed to characterize a patient with high levels of activation.

Digital Tools to Support SDM

To support SDM while using the continuous development and use of digital solutions, researchers have started to investigate how digital interventions may support SDM. Digital mental health interventions, such as interventions including a smartphone app, have been found to significantly outperform control groups [10]. However, the evidence on digital mental health interventions to support SDM is sparse, but a recent meta-analysis found that digital SDM interventions may have an effect on patient activation, decisional conflict, working alliance, and severity of general symptoms [11]. The meta-analysis also concluded that while digital interventions to support SDM are promising, the limited evidence is in need for quality research.

This study aimed to provide new evidence on the effectiveness of a digital SDM intervention in mental health care and strengthen the evidence on how digital tools may promote patient activation. We evaluated the effectiveness of a digital solution to support SDM in an outpatient setting for people diagnosed with schizophrenia. We hypothesized that the intervention would support SDM, resulting in higher levels of self-perceived patient activation. With higher levels of patient activation, we also expected to see improvements in working alliance, hope, self-efficacy, satisfaction, feeling prepared for decision-making, confidence in communicating with one's provider, severity of symptoms, level of functioning, number of hospitalizations, and adherence.

Methods

Trial Design and Setting

This study was a 2-arm, assessor-blinded, randomized parallel-group trial conducted in 9 outpatient treatment sites called OPUS in the Capital Region of Denmark. OPUS is a 2-year treatment program providing specialized early intervention treatment to patients with a debuting diagnosis of schizophrenia or related psychotic disorders in the age group of 18 to 35 years in Denmark. This trial compared a control group receiving treatment as usual (TAU) with an intervention group receiving a smartphone app as a supplement to TAU. The participants were recruited between January 2019 and March 2021. Assessments were conducted at baseline, 3 months after baseline (midintervention time point), and 6 months after baseline (postintervention time point). Detailed information on the trial design and methodology of the study is available in the study protocol [12].

Participants and Eligibility Criteria

Eligible patients were referred to the study by their primary providers. Patients were eligible for inclusion if they were receiving treatment in OPUS (see the section *Treatment as Usual* for information on OPUS), had at least 6 months left of their OPUS program, access to a smartphone, and understood Danish. Patients were enrolled after meeting a staff member from the research team who provided detailed verbal and written information about the study, and written consent was obtained.

Randomization and Blinding

Participants were randomized with an even allocation of 1:1 to either the intervention group (TAU plus app) or the control group (TAU minus app). Randomization was performed after completion of the baseline assessment. Block randomization was used to achieve balance in the allocation of participants to both treatment arms. The block sizes were randomly altered among 2, 4, and 6. The block sizes were concealed from the researchers during recruitment. The nonstratified randomization sequence was computerized and facilitated by the Odense Patient Data Explorative Network (OPEN) to ensure allocation concealment. The concealment was kept digital at OPEN until data collection ended and data analysis began. To ensure blinding of the data analyst, OPEN provided information on which group participants had been part of but without labeling the groups. This way the data analysis could be performed without bias by knowing who had been in the control group and who had been in the intervention group. After the whole research group had accepted the results of the data analysis and conclusions had been drawn, OPEN was contacted to reveal the labeling of the 2 groups.

Researchers collecting and analyzing data were blinded, but given the nature of the intervention, patients and health care providers were not blinded. All patients were at each visit, with the researcher thoroughly instructed not to mention anything about their randomization allocation. Therefore, all questionnaire outcomes (answered by the patient or provider) were not blinded, whereas the interview outcomes (assessor-rated) were blinded.

Interventions

Treatment as Usual

Participants randomized to the control group continued with TAU and did not receive the digital SDM intervention. TAU in this trial was provided by OPUS, a treatment facility offering specialized early intervention by combining three key elements: (1) assertive community treatment aimed at maintaining or developing the patient's coping skills and integration in society; (2) family involvement through multifamily groups and single-family sessions; and (3) social skills training to support patients with impaired social skills [13]. Patients starting OPUS are assigned to a primary provider with weekly sessions (excluding group sessions) lasting for approximately 40 to 60 minutes. Primary providers in OPUS may have a background as a psychologist, nurse, social worker, physiotherapist or vocational therapist. The OPUS treatment does facilitate recovery and SDM elements with its patient-centered approach where patients are considered *a long-awaited guest who should feel at home during a visit* and who are encouraged to take an active part in the treatment. Nevertheless, we chose to conduct the study in OPUS because the results from our pilot study indicated that younger adults with schizophrenia spectrum disorders showed a positive attitude toward using a digital tool to support their care [14]. During recruitment for the trial, the providers had approximately 15 patients at a time and were able to have patients in both groups.

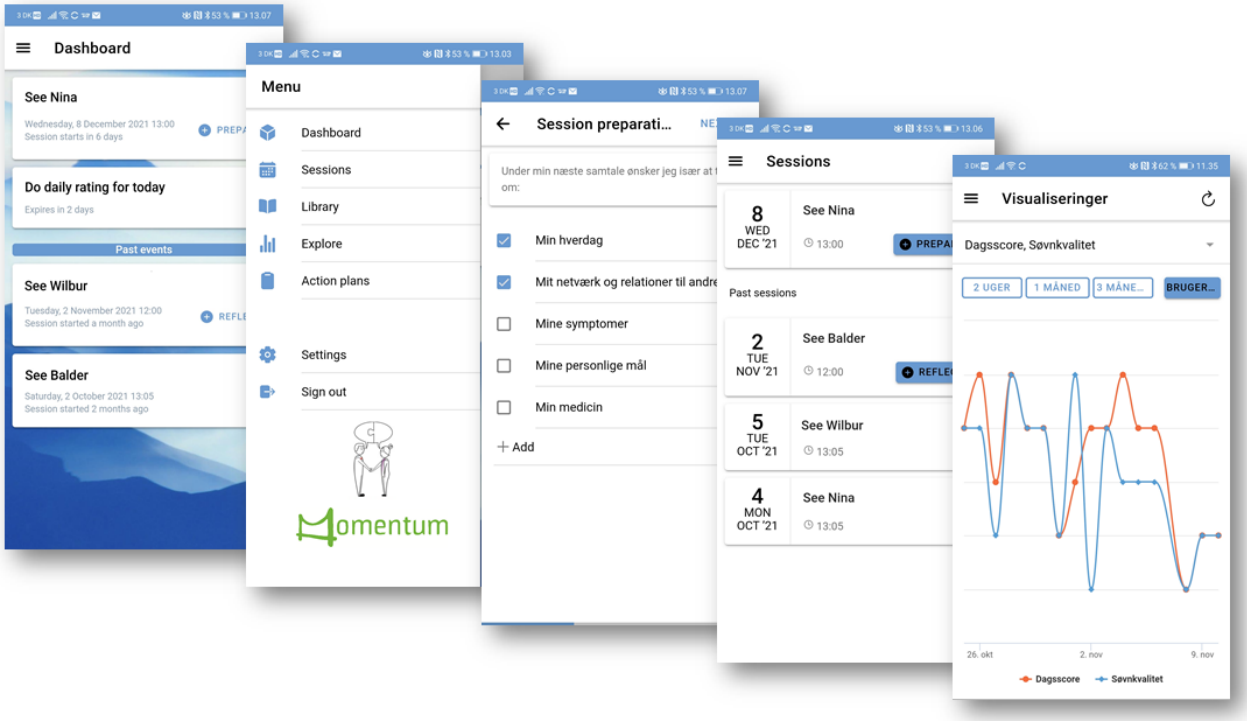
The study protocol provides more information on TAU [12].

The Intervention Group

Before the trial, we developed a digital SDM tool for the process of cocreation among patients, providers, and researchers, with preparation for treatment consultation as the main function. A pilot study revealed that the tool was perceived to be useful with relevant content by patients and providers [14]. On the basis of feedback from the pilot study, the app was adjusted accordingly and included a new functionality, an option to perform a daily self-assessment.

The intervention group continued with TAU and was invited to use the digital system provided by the IT company Monsenso. The digital SDM tool tested in this trial consisted of a smartphone app for the patient with functions, such as preparation for consultation, daily self-assessments, action plans, and educational material. The app was synchronized to a web portal that the patient's provider could access before the consultation. The intention was that the patient could use the app outside of the consultation and that the provider before an upcoming consultation could become aware of what the patient would like to address at the consultation while also seeing how the patient had scored themselves on the self-assessments. The patients were encouraged to use the app daily or what felt meaningful. Patients were shown how to set up reminders within the app to enable push messages. Enabling these push messages was voluntary. Providers were encouraged to use the web portal before a consultation; however, there were no options for reminders or push messages for the providers. Most importantly, patients and providers were encouraged to discuss how best to use the system and how to incorporate it to support the consultations. The digital system is illustrated in Figure 1.

Figure 1. Digital shared decision-making tool for smartphones.



Outcomes

Baseline Parameters

Information on the characteristics of both patients and providers was collected. As preferences in clinical decision-making have been found to be related to patient involvement, the Clinical Decision Making Style (CDMS) questionnaire was completed by both the patient and provider at baseline. The questionnaire consisted of 2 subscales: preference for participation in decision-making and preference for receiving information. This questionnaire was only completed at baseline because research indicates that CDMS scores are stable over 12 months [15].

Primary Outcome

Our primary outcome was the difference in self-perceived patient activation between the groups at the postintervention time point, as measured by the 10-item Consumer Health Activation Index for mental health (CHAI-MH) [16].

Secondary Outcomes

Our secondary outcomes consisted of questionnaires completed by participating patients and providers, a clinical interview, and data from the Danish National Patient Register. Patients completed the following questionnaires: self-perceived feeling of hope and optimism measured by the 6-item Adult State Hope Scale [17], self-efficacy measured by the 10-item General Self-Efficacy Scale (GSE) [18], confidence in communicating with one's provider measured by the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI) Questionnaire [19], therapeutic alliance between the patient and provider measured by the 12-item Working Alliance Inventory–short form (WAI-S) [20], feeling prepared to make a treatment decision by the 10-item Preparation for Decision-Making (PrepDM) [21], and satisfaction with treatment measured by the 8-item Client Satisfaction Questionnaire (CSQ) [22]. In addition, a clinical interview was conducted to assess the participants' positive and negative symptoms, together with their level of functioning. We used the Scale for the Assessment of Positive Symptoms (SAPS) [23], Scale for the Assessment of Negative Symptoms (SANS) [23], Global Assessment of Functioning (GAF) [24] and Personal and Social Performance Scale (PSP) [25]. A blinded researcher conducted the interviews. Providers completed 2 questionnaires for each of their patients participating in the trial: the therapeutic alliance between the provider and patient measured by the 12-item WAI–S [20] and the patient's engagement measured by the Service Engagement Scale (SES)—collaboration subscale [26]. Finally, we collected data for all participating patients from the Danish National Patient Register-Psychiatry on the following: number of hospital admissions, length of admissions in days, and adherence to OPUS appointments. Reasoning for choosing the outcomes can be found in the study protocol.

Explorative Outcomes

To explore the acceptance and perceived usefulness of the smartphone app, participants in the intervention group completed the 4-item App Rating Questionnaire and the 4-item Mobile App Rating Scale—subscale subjective quality rating at the postintervention time point [27,28]. In addition, objective data on the use of the system (user sessions per day, screen views per day, screens per session, session duration and session instances, and user retention) were provided by Monsenso.

Sample Size

As stated in our protocol, a sample size of 180 participants was estimated to be needed to detect a significant difference between the intervention and control groups, with an effect size of 0.42 on the CHAI-MH scale. The effect size was calculated based

on previous research that measured patient activation, as described in the research protocol. For both the primary and secondary outcomes, a power of 80% and an α of .05 was chosen to reject the null hypothesis that the population means of the 2 groups are equal. Before recruitment, we estimated that 30% would be lost to follow-up (ie, not responding to contact at the postintervention time point). To adjust for this, a sample size of 260 participants is needed. However, during the recruitment of the first 100 participants, only 7 (7%) were lost to follow-up. As the rate was significantly lower than anticipated, we changed our estimated percentage of lost-to-follow-up from 30% to 7%, resulting in a required sample size of 194 participants.

Statistical Methods

For the statistical analysis, the principles of intention to treat (ITT) were followed with a 2-tailed level of significance for all statistical tests set at .05. Analyses were performed using SAS Enterprise Guide 7.1. Differences in patient characteristics between the 2 groups were assessed using the 2-sample *t* test (2-tailed), chi-square test, and Fisher exact test (for variables with <5 observations). Generalized linear mixed effects regression analyses were performed to assess the 6-month intervention. A binary logistic regression was performed to evaluate the impact of the intervention on participants' use of antipsychotic medication. Negative binomial regression was performed for count outcomes to estimate the incidence rate ratios on the number of hospitalizations and the length of hospitalization at the postintervention time point based on the group allocation. To handle missing data, we created and analyzed 100 imputed data sets using multiple imputations by chained equations using the group variable (*intervention* and *control*); the use of antipsychotic medicine variable; completed interview at the postintervention time point; and the participants' baseline, midintervention, and postintervention scores. The use of antipsychotic medication at baseline (score=yes or no) was used as a variable for the imputed data sets due to a significant difference between groups in the use of antipsychotic medication at baseline. During data analyses, we found that participants who completed the postintervention interview scored lower on the CHAI-MH than participants who had not completed the interview. Although there were no between-group differences, we decided to include this dichotomous variable when computing multiple imputations for the questionnaire outcomes. For the imputed data sets on the interview outcomes, we did not use this variable because its value would be the same for all imputed data. Outcome scales with partially missing values were regarded as completely missing. For each outcome, an estimate of the effect was calculated for each imputed data set and finally combined using the Rubin rule. We also performed a complete case analysis for comparative purposes. The midintervention assessment was included for explorative purposes to assess whether a potential effect occurred before or after 3 months of intervention.

Ethical Considerations

The trial was approved by the Regional Ethics Committee in the Capital Region of Denmark under file number H-17025550 and the Knowledge Centre on Data Protection Compliance (Videnscenter for Dataanmeldelser) under approval number P-2019-502. The trial was registered at ClinicalTrials.gov under the identifier NCT03554655. No economic compensation was provided for participation.

Changes From the Protocol

As stated in our study protocol, we were interested in evaluating the mean duration per session for which the participants used the smartphone app. However, due to technical limitations, we were unable to assess the duration for which the participants used the app.

Owing to the COVID-19 pandemic, several assessments were conducted on the web; however, no statistically significant differences in scores for participants being assessed physically or virtually were detected.

Owing to a fire accident at OVHcloud (a global cloud service provider that stores Monsenso's data), the digital system became unavailable for approximately 1 month during which participants were unable to access the app and web portal. This downtime affected approximately 36 participants in the intervention group. Owing to blinding, the research group did not directly reach out to participants. Instead, all providers were contacted regarding this issue and instructed to inform participants of the system being unavailable in the intervention group. After the system became available again, providers were instructed to inform participants to use the app again. In addition to the accident, a failure in the Monsenso back-up system resulted in a loss of data for the last month leading up to the fire accident. To assess whether the interruption had an impact on the use of the system, objective data on the use of the system before the accident were compared with data on the use after the system became available again.

In our study protocol, we calculated Cohen κ for the CDMS questionnaire to assess the level of agreement between patients and providers. However, due to the data structure of the CDMS, this was not possible, and we instead performed a *t* test to assess if there was a statistically significant difference between the responses of patients and providers.

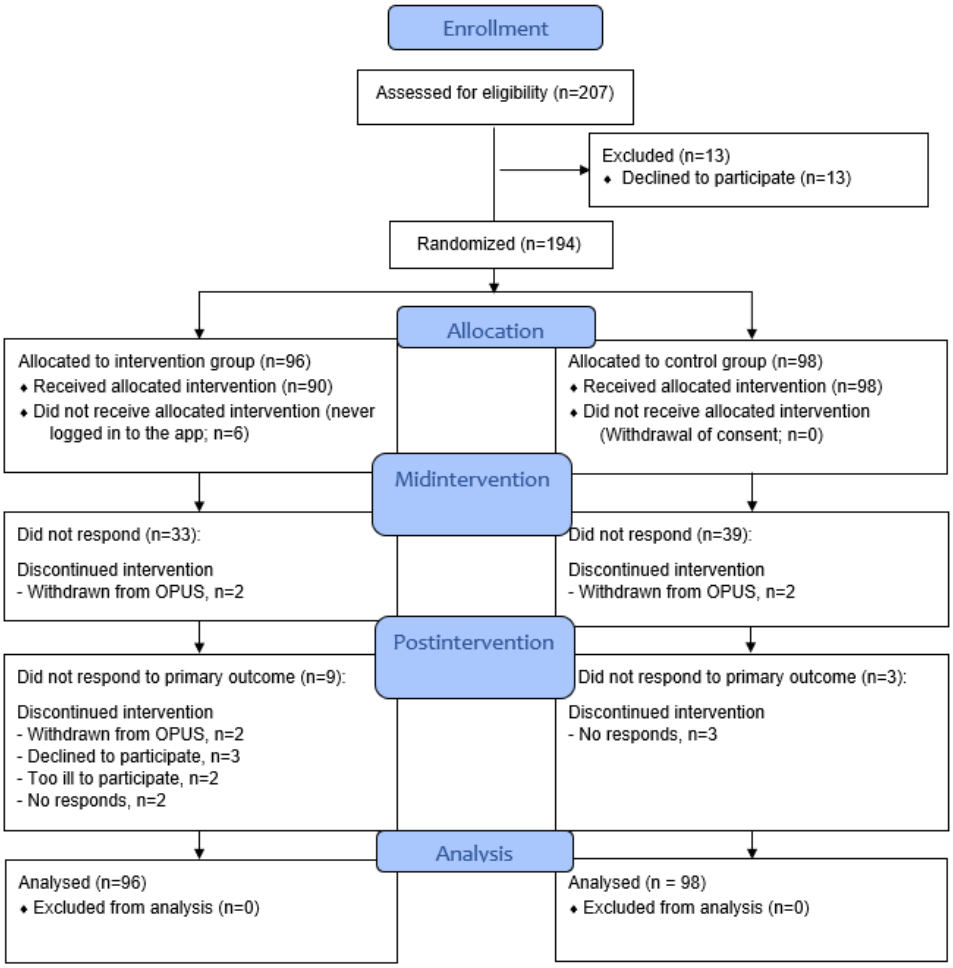
Although we planned to assess the effect of the intervention based on duration in OPUS (eg, patients at the beginning of their treatment versus those at the end of their treatment), we were unable to do so because of safety procedures regarding merging patient-reported outcome data with data from the Danish National Health Registers.

Results

Overview

Figure 2 presents the CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the participants in the Momentum Trial. In total, 194 participants were included and randomized, with 98 to the control group and 96 to the intervention group. Recruitment began in January 2019, and the last patient was enrolled in September 2021.

Figure 2. Flow diagram for the Momentum Trial.



Background Characteristics

Tables 1 and 2 present the sociodemographics and background characteristics of patients and providers, respectively. The intervention and control groups differed in terms of age and use of antipsychotic medication, while no between-group differences were observed in gender, diagnosis, relationship status, level of education, employment status, duration of received treatment in OPUS at baseline, or scores on the CDMS questionnaire.

We also assessed the level of agreement between patients and providers on the WAI-S and CDMS. We observed a significant difference between the patients and providers on the information subscale (mean difference 0.33, 95% CI 0.18-0.46; *P*<.001) and on the decision-making subscale (mean difference −0.20, 95% CI −0.27 to −0.12; *P*<.001). While the mean differences are rather small, the results suggest that patients have a higher desire to be provided with information than providers' desire to provide information. In contrast, patients have a smaller desire to be active participants in decision-making compared with the providers' desire for active participation from the patient.

For the WAI-S scale, we considered an agreement between the patient and provider if their scores were within 12 points. With this range of agreement, the weighted Cohen κ was calculated to be 0.45 at baseline and 0.43 at the postintervention time point, indicating a stable yet weak level of agreement between patients and their provider on the working alliance.

There were no between-group differences on the WAI-S or the CDMS questionnaires.

Table 1. Sociodemographics and background characteristics of patients.

	Control group (n=98)	Intervention group (n=96)	Overall (N=194)
Age (years), mean (SD)	24.3 (4.3)	22.7 (3.7)	23.5 (4.1)
Gender, n (%)			
Woman	55 (56.1)	65 (67.7)	120 (61.9)
Man	39 (39.8)	26 (27.1)	65 (33.5)
Nonbinary	4 (4.1)	5 (5.2)	9 (4.6)
Diagnosis, n (%)			
Schizophrenia	40 (40.8)	28 (29.2)	68 (35.1)
Schizotypal	48 (49.0)	54 (56.3)	102 (52.6)
Other nonorganic psychosis	8 (8.2)	14 (14.6)	22 (11.3)
Schizoaffective	2 (2.0)	0 (0)	2 (1.0)
In a relationship, n (%)			
Not in a relationship	54 (55.1)	46 (47.9)	100 (51.5)
In a relationship	44 (44.9)	50 (52.1)	94 (48.5)
Level of education, n (%)			
Primary school not completed	2 (2.0)	1 (1.0)	3 (1.5)
Primary school completed	29 (29.6)	34 (35.4)	63 (32.5)
High school or higher completed	67 (68.4)	61 (63.5)	128 (66)
Employment status, n (%)			
Employed	15 (15.3)	11 (11.5)	26 (13.4)
Student	34 (34.7)	47 (49.0)	81 (41.8)
Unemployed and not a student	49 (50.0)	38 (39.6)	87 (44.8)
Use of antipsychotics, n (%)			
Yes	76 (77.6)	57 (59.4)	133 (68.6)
No	22 (22.4)	39 (40.6)	61 (31.4)
Clinical Decision Making Style questionnaire score, mean (SD)			
Information	3.3 (0.5)	3.3 (0.6)	3.3 (0.5)
Participation in decision-making	2.1 (0.3)	2.1 (0.3)	2.1 (0.3)
Duration of received treatment in OPUS at baseline (days), mean (SD)	306.52 (150.41)	262.10 (149.84)	284.54 (151.38)

Table 2. Sociodemographics and background characteristics of providers.

	Overall (n=76)
Age (years), mean (SD)	43.1 (10.1)
Gender, n (%)	
Woman	66 (87.0)
Man	10 (13.0)
Experience (years), mean (SD)	13 (8.2)
Education of provider, n (%)	
Nurse	30 (39.5)
Occupational therapist	15 (19.7)
Psychologist	13 (17.1)
Social worker	11 (14.5)
Pedagogue	2 (2.6)
Other	4 (5.3)
Missing	1 (1.3)
Clinical Decision Making Style questionnaire score, mean (SD)	
Information	3.0 (0.5)
Participation in decision-making	2.3 (0.3)

Use of Intervention

On the basis of objective data from Monsenso, only 86 participants used the app, meaning that 10 never started using the app despite being invited to use it. Although the reasons for this were not explored, there were reports of technical limitations where the participants' phones did not support the app.

Owing to the fire accident at OVHcloud, we encountered a month in which the app was not accessible. To evaluate the impact of a "pause" in the intervention, we compared app use for the last 3 months leading up to the fire accident with data for the 3 months after the system became available again. We observed that 47% (17/36) of active users did not log back into the app, and there was a decrease in the number of app sessions from 1260 to 491 (61%), indicating that the fire accident had an impact on the participants' use of the app.

Lost-to-Follow-up

A total of 8.2% (16/194) of participants did not participate at the postintervention time point (11 participants from the intervention group and 5 from the control group). The difference in lost-to-follow-up between groups was mainly due to 4 participants in the intervention group who ended their OPUS treatment prematurely and 3 participants who refused to participate at the postintervention time point (vs 2 and 0 participants in the control group, respectively). The most frequent reason for loss to follow-up in the control group was not responding to contact (3 participants). According to our power calculation, we required 180 participants who completed the baseline and postintervention assessments to reach an adequate level of power of 0.42. We enrolled 178 participants who completed both baseline and postintervention assessments; therefore, we did not reach the level of power we had aimed at.

There was a large discrepancy between the completed questionnaires and the completed interviews at the postintervention time point. In the intervention group, 89% (85/96) completed at least 1 questionnaire, while 56% (54/96) completed the interview. In the control group, 95% (93/98) pf participants completed at least 1 questionnaire, while 70% (69/98) completed the interview.

The percentage of missing values across the 11 outcome measurements for each participant varied from 0% to 3% at baseline, and from 8% to 36% at the end of the intervention. In total, 52.4% of the records were incomplete, meaning that they had ≥1 missing variables at baseline or after the intervention. The variables with the highest proportion of missing information when events were combined were clinical interview data (SAPS, SANS, GAF, and PSP), where approximately 46% were missing. For the questionnaire variables, the highest proportion of missing data was found for the CSQ and PrepDM (approximately 40% missing data).

Intention to Treat

Tables 3 and 4 show the results of the mid- and postintervention ITT analyses, while Figure 3 illustrates the effect of the intervention. The Momentum Trial resulted in a statistically significant difference between the intervention and control groups in our primary outcome, CHAI-MH (mean difference 4.39, 95% CI 0.99-7.79; Cohen *d*=0.33; *P*=.01), favoring the intervention group. For the secondary outcomes, there were 2 scales with a minor statistically significant difference: PEPPI (mean difference 1.85, 95% CI 0.01-3.69; Cohen *d*=0.24; *P*=.05) and PrepDM (mean difference 5.12, 95% CI 0.16-10.08; Cohen *d*=0.27; *P*=.04), both favoring the intervention group. For the remaining outcome we found no statistically significant differences between the groups: Hope (mean difference 1.66, 95% CI −0.44 to 3.75; Cohen *d*=0.20; *P*=.12), GSE (mean

difference 1.12, 95% CI −0.32 to 2.57; Cohen *d*=0.19; *P*=.13), WAI-S (mean difference 2.43, 95% CI −0.25 to 5.12; Cohen *d*=0.22; *P*=.08), CSQ (mean difference 0.89, 95% CI −0.13 to 1.91; Cohen *d*=0.22; *P*=.09), SAPS-Psychotic (mean difference −0.2, 95% CI −0.43 to 0.04, Cohen *d*=−0.20, *P*=.10), SANS (mean difference −0.14, 95% CI −0.33 to 0.04; Cohen *d*=−0.18; *P*=.13), SAPS-Disorganized (mean difference −0.02, 95% CI −0.16 to 0.11; Cohen *d*=−0.06; *P*=.71), GAF (mean difference 1.35, 95% CI −1.01 to 3.72; Cohen *d*=0.13; *P*=.26), PSP (mean difference 1.38, 95% CI −0.68 to 3.44; Cohen *d*=0.13; *P*=.19). There were no statistically significant differences between provider scores WAI-S Provider (mean difference −0.81, 95% CI −2.5, 0.87; Cohen *d*=−0.09; *P*=.34) or SES (MD=−0.10, 95% CI −0.48 to 0.28; Cohen *d*=−0.06; *P*=.60). Finally, we found no statistically significant difference between the intervention and control groups in the use of antipsychotic medication at the postintervention time point (odds ratio 0.46, 95% CI 0.13-1.61; *P*=.23).

Data from the Danish National Patient Register revealed no significant differences between the intervention and control groups in the mean number of hospitalizations (incidence rate ratio 0.80, 95% CI 0.27-2.37; *P*=.69) or length of admission in days (incidence rate ratio 0.76, 95% CI 0.11-5.53, *P*=.79). The incidence rate of hospitalization for the intervention group was 0.11 (95% CI 0.05-0.25), while that for the control group was 0.14 (95% CI 0.07-0.30). The incidence rate of days hospitalized for the intervention group was 1.60 (95% CI 0.39-6.56), while that of the control group was 2.10 (95% CI 0.52-8.46)

In terms of contacts to OPUS (eg, consultations), the intervention group had 2572 contacts (4.47 contacts per person per month) versus the control group having 2694 contacts (4.58 contacts per person per month), a nonsignificant difference.

Table 3. Intention-to-treat analyses of primary and secondary outcomes.

Intention-to-treat analyses	Intervention group				Control group				*P* value[a]	Cohen *d*
	Baseline, mean (SD)	Midintervention (3 months), mean (SD)	Postintervention (6 months), mean (SD)	Value, n (%)	Baseline, mean (SD)	Midintervention (3 months), mean (SD)	Postintervention (6 months), mean (SD)	Value, n (%)		
CHAI-MH[b]	55.52 (13.20)	61.04 (12.71)	64.91 (13.42)	96 (100)	56.49 (13.66)	61.33 (12.59)	61.19 (13.5)	98 (100)	.01	0.33
PEPPI[c]	34.45 (8.40)	35.56 (8.85)	38.95 (7.13)	96 (100)	34.93 (8.7)	36.54 (7.36)	37.36 (8.1)	98 (100)	.05	0.24
Hope	26.04 (9.04)	28.53 (6.99)	31.80 (7.36)	96 (100)	26.73 (9.12)	29.94 (7.71)	30.34 (9.29)	98 (100)	.12	0.20
GSE[d]	23.02 (5.44)	24.47 (5.74)	26.74 (6.08)	96 (100)	23.79 (5.96)	25.70 (4.63)	26.10 (5.55)	98 (100)	.13	0.19
WAI-S[e]	66.08 (10.78)	—[f]	69.21 (10.28)	96 (100)	66.56 (11.51)	—	67.20 (11.88)	98 (100)	.08	0.22
PrepDM	58.33 (18.43)	—	66.84 (19.15)	96 (100)	62.57 (18.07)	—	64.58 (18.65)	98 (100)	.04	0.27
CSQ[g]	26.27 (3.54)	—	27.34 (3.83)	96 (100)	26.63 (3.99)	—	26.76 (4.27)	98 (100)	.09	0.22
Psychotic dimension[h]	2.06 (1.11)	—	1.42 (0.82)	96 (100)	2.02 (1.19)	—	1.59 (1.08)	98 (100)	.10	−0.20
Negative dimension[i]	1.82 (0.92)	—	1.36 (0.81)	96 (100)	1.83 (0.93)	—	1.51 (0.77)	98 (100)	.13	−0.18
Disorganized dimension[j]	0.53 (0.56)	—	0.31 (0.38)	96 (100)	0.62 (0.67)	—	0.38 (0.44)	98 (100)	.71	−0.06
GAF[k]	56.14 (12.52)	—	62.39 (11.06)	96 (100)	53.84 (12.05)	—	59.34 (10.34)	98 (100)	.26	0.13
PSP[l]	57.42 (12.23)	—	63.20 (10.78)	96 (100)	55.36 (11.76)	—	60.25 (10.10)	98 (100)	.19	0.13
WAI-S (P)[m]	63.51 (9.30)	—	64.11 (9.30)	96 (100)	64.35 (8.44)	—	65.71 (8.25)	98 (100)	.34	−0.09
SES[n]	2.36 (1.91)	—	2.34 (1.82)	96 (100)	2.00 (1.65)	—	1.96 (1.71)	98 (100)	.60	−0.06

[a]Comparison of means between intervention group and control group postintervention.

[b]CHAI-MH: Consumer Health Activation Health Index–mental health version.

[c]PEPPI: Perceived Efficacy in Patient-Physician Interactions.

[d]GSE: General Self-Efficacy.

[e]WAI-S: Working Alliance Inventory–Short.

[f]Not available.

[g]CSQ: Client Satisfaction Questionnaire.

[h]Global item scores of hallucinations and delusions.

[i]Global item scores of affective flattering, alogia, avolition-apathy, and anhedonia.

[j]Global item scores of bizarre behaviors, formal thought disorder and single item score of inappropriate affect.

[k]GAF: Global Assessment of Functioning.

[l]PSP: Personal and Social Performance Scale.

[m]WAI-S (P): Working Alliance Inventory–Short (Provider version).

[n]SES: Service Engagement Scale.

Table 4. Intention-to-treat analyses of hospital admissions and use of medication.

Intention-to-treat analyses	Intervention group				Control group				*P* value[a]
	IRR[b]	IR[c]	Odds ratio (95% CI)	Value, n (%)	IRR	IR	Odds ratio (95% CI)	Value, n (%)	
Number of hospital admissions	0.80 (0.27-2.37)	0.11 (0.05-0.25)	—[d]	96 (100)	1 (reference)	0.14 (0.07-0.30)	—	98 (100)	.69
Number of days admitted	0.76 (0.11-5.53)	1.60 (0.39-6.56)	—	96 (100)	1 (reference)	2.10 (0.52-8.46)	—	98 (100)	.79
Use of medication	—	—	0.46 (0.13-1.61)	96 (100)	—	—	1 (reference)	98 (100)	.23

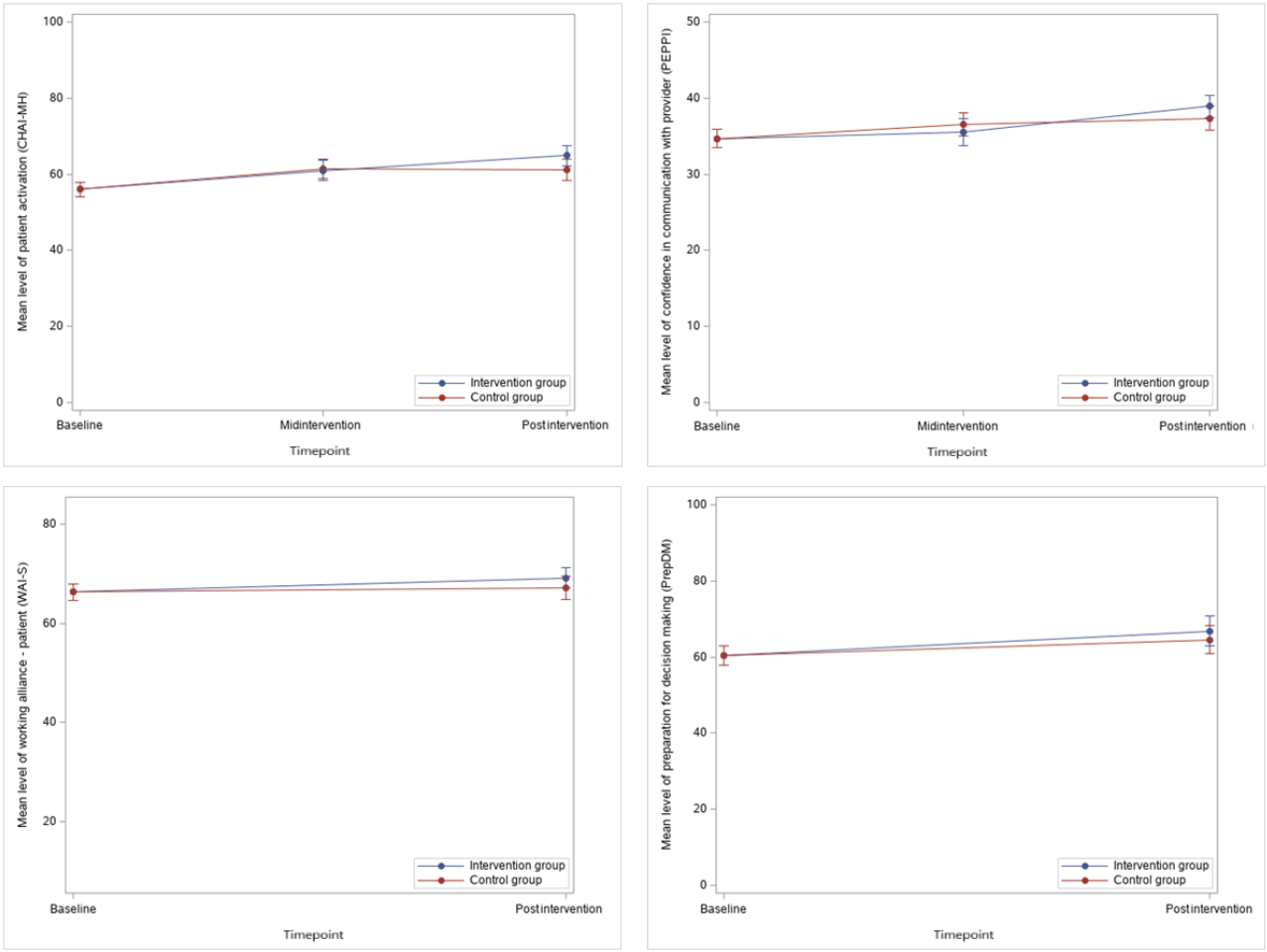
[a]Comparison of means between intervention group and control group postintervention.

[b]IR: incidence rate.

[c]IRR: incidence rate ratio.

[d]Not available.

Figure 3. Mean scores on the primary outcome (Consumer Health Activation Index for mental health [CHAI-MH]) and selected secondary outcomes; Perceived Efficacy in Patient-Physician Interactions (PEPPI), PrepDM, and Working Alliance Inventory–short (WAI-S) for intervention and control groups over time with 95% CIs.



Complete Case Analyses

Table 5 presents the results of the complete case analyses. The analyses showed similar results, although with larger variation, as in the ITT analyses; however, the statistically significant difference between groups on PEPPI was no longer present (mean difference 1.80, 95% CI −0.05 to 3.64, *P*=.06).

Table 5. Complete case analyses of primary and secondary outcomes.

Complete case analyses	Intervention group						Control group						*P* value[a]	Cohen *d*
	Baseline		Midintervention (3 months)		Postintervention (6 months)		Baseline		Midintervention (3 months)		Postintervention (6 months)			
	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)		
CHAI-MH[b]	55.54 (13.27)	95 (98.96)	60.60 (13.45)	63 (65.63)	65.02 (13.88)	84 (87.5)	56.49 (13.66)	98 (100)	60.31 (13.67)	59 (60.20)	61.16 (13.77)	93 (94.90)	.01	0.31
PEPPI[c]	34.49 (8.49)	94 (97.92)	35.82 (9.45)	61 (63.54)	39.07 (7.43)	82 (85.42)	35.01 (8.70)	97 (98.98)	36.76 (7.89)	58 (59.18)	37.17 (8.26)	92 (93.88)	.06	0.23
Hope	26.04 (9.13)	94 (97.92)	28.50 (7.82)	60 (62.50)	31.87 (7.59)	84 (87.50)	26.74 (9.17)	97 (98.98)	28.95 (8.52)	57 (58.16)	30.33 (9.46)	91 (92.86)	.16	0.17
GSE[d]	23.01 (5.51)	92 (95.83)	24.22 (6.13)	58 (60.42)	26.70 (6.56)	73 (76.04)	23.79 (5.96)	98 (100)	25.18 (5.25)	57 (58.16)	25.87 (5.67)	87 (88.78)	.12	0.19
WAI-S[e]	66.08 (11.02)	91 (94.79)	—[f]	—	69.36 (10.53)	76 (79.17)	66.53 (11.67)	95 (96.94)	—	—	66.80 (12.42)	80 (81.63)	.09	0.20
PrepDM	58.42 (18.81)	92 (95.83)	—	—	66.93 (20.37)	79 (82.29)	62.66 (18.15)	96 (97.96)	—	—	64.79 (19.47)	85 (86.73)	.04	0.26
CSQ[g]	26.36 (3.60)	91 (94.79)	—	—	27.36 (3.99)	74 (77.08)	26.63 (3.99)	98 (100)	—	—	26.76 (4.52)	84 (85.71)	.09	0.20
Psychotic dimension[h]	2.06 (1.11)	96 (100)	—	—	1.40 (0.94)	54 (56.25)	2.02 (1.19)	98 (100)	—	—	1.59 (1.12)	69 (70.41)	.15	−0.17
Negative dimension[i]	1.82 (0.92)	96 (100)	—	—	1.17 (0.86)	53 (55.21)	1.83 (0.93)	98 (100)	—	—	1.51 (0.85)	65 (66.33)	.07	−0.20
Disorganized dimension[j]	0.53 (0.56)	96 (100)	—	—	0.30 (0.43)	53 (55.21)	0.62 (0.67)	98 (100)	—	—	0.38 (0.46)	65 (66.33)	.49	−0.10
GAF[k]	56.14 (12.52)	96 (100)	—	—	65.80 (11.35)	54 (56.25)	53.84 (12.05)	98 (100)	—	—	59.88 (11.17)	69 (70.41)	.20	0.13
PSP[l]	57.42 (12.23)	96 (100)	—	—	66.56 (10.43)	54 (56.25)	55.36 (11.76)	98 (100)	—	—	60.78 (10.73)	69 (70.41)	.14	0.14
WAI-S (P)[m]	63.55 (9.48)	92 (95.83)	—	—	64.09 (9.55)	90 (93.75)	64.34 (8.71)	92 (93.88)	—	—	65.88 (8.53)	86 (87.76)	.25	−0.11
SES[n]	2.35 (1.94)	93 (96.88)	—	—	2.32 (1.86)	91 (94.79)	2.00 (1.70)	92 (93.88)	—	—	1.98 (1.80)	86 (87.76)	.55	0.06
Use of medication[o]	—	96 (100)	—	—	—	54 (56.25)	—	98 (100)	—	—	—	69 (70.41)	.09	—

[a]Comparison of means between intervention group and control group postintervention.

[b]CHAI-MH: Consumer Health Activation Health Index–mental health version.

[c]PEPPI: Perceived Efficacy in Patient-Physician Interactions.

[d]GSE: General Self-Efficacy.

[e]WAI-S: Working Alliance Inventory–Short.

[f]Not available.

[g]CSQ: Client Satisfaction Questionnaire.

[h]Global item scores of hallucinations and delusions.

[i]Global item scores of affective flattering, alogia, avolition-apathy, and anhedonia.

[j]Global item scores of bizarre behaviors, formal thought disorder and single item score of inappropriate affect.

[k]GAF: Global Assessment of Functioning.

[l]PSP: Personal and Social Performance Scale.

[m]WAI-S (P): Working Alliance Inventory–Short (Provider version).

[n]SES: Service Engagement Scale.

[o]There was no significant difference (*P*=.09) in the odds of using antipsychotic medication at the postintervention time point between the intervention

group compared to the control group (OR 0.36, 95% CI 0.11-1.17).

Explorative Outcomes

Although the aim of the study was to assess the effectiveness of the intervention after 6 months, we also assessed its effectiveness after 3 months for selected outcomes to explore when a potential effect would occur. On the basis of the ITT analyses and complete case analyses, we found no statistically significant differences between baseline and midintervention on CHAI-MH, PEPPI, Hope, or GSE.

Objective data on use of the app revealed that the intervention group had a mean use of 0.55 log-ins per day during their active use period (corresponding to roughly one session every second day). The active use period ranged from 1 day to 180 days, with a mean of 39 (SD 37.70) days, whereas the mean number of unique sessions was 23 ranging from 1 session to 148 sessions. When using the app, participants saw an average of 20 different screens, ranging from 5 to 28 screen views. Finally, 55% (47/96) of participants in the intervention group logged in after the first month. On the basis of the App Rating Questionnaire, participants were somewhat satisfied with the app (mean score was 6.36 out of 12), while they rated the app to be of average quality (mean score was 2.85 out of 5).

Discussion

Principal Findings

This study presents the results of a clinical trial investigating a digital SDM tool to promote patient activation for people diagnosed with schizophrenia. The study found a statistically significant difference in our primary outcome, patient activation, CHAI-MH (mean difference 4.39, 95% CI 0.99-7.79; Cohen *d*=0.33; *P*=.01), favoring the intervention group. These findings confirm our hypothesis that a digital SDM tool may promote patient activation by supporting the collaborative process between patients and their providers and is in concordance with recent meta-analyses on the effectiveness of digital SDM interventions that found these types of interventions to have an effect on patient activation [11]. The effect size (Cohen *d*) for patient activation was 0.33, which may be interpreted as a small effect size. According to our protocol and power calculations, we expected to find an effect size of 0.42, thereby not reaching the anticipated effect. In addition, it is unclear whether this effect size is clinically relevant. In somatic care, patient activation has been found to play an important role in improving quality and health outcomes, where every 10 points in patient activation were associated with a 1% decreased probability of visiting an emergency unit, being obese, or smoking [29]. The mean difference in our trial was 4.39 and somewhat far off the 10 points found in the study by Greene and Hibbard [29]. However, such studies have not been conducted in mental health care and are needed to better assess the minimal clinical relevance of people with a mental health condition having higher levels of patient activation. However, as argued in the trial by Hamann et al [2], SDM interventions can improve the feeling of being involved in one's treatment, which may be particularly useful for people feeling involuntarily treated or those who refuse treatment due to a lack of insight in their care.

Our intervention also had an effect on 2 secondary outcomes: PEPPI and PrepDM. Although these results were close to the 0.05 cut-off level, they favored the intervention group, similar to our primary outcome. In addition, for the complete case analyses, we found a statistically significant difference in PrepDM, favoring the intervention group. Thus, our trial indicates that a digital SDM tool is effective in improving patient activation, feeling prepared for decision-making, and confidence in communicating with one's provider.

Although none of the other secondary outcomes had a statistically significant effect, most secondary outcomes favored the intervention group. One unexpected finding was that we did not find a statistically significant effect on satisfaction since SDM has been strongly advocated as a process to increase patient satisfaction with treatment. However, similar to other trials, we encountered a ceiling effect on the CSQ scale, with 44.8% (87/194) of participants scoring ≥29 out of 32 [30].

Despite the difference in self-perceived patient activation, we found no difference in how providers perceived their patients' level of engagement via the SES (mean difference −0.10, 95% CI −0.48 to 0.28; Cohen *d*=−0.06; *P*=.60). It may be intuitive to assume that increased levels of patient activation are associated with an increase in providers' perceptions of patients' level of engagement. However, studies have found that some providers find it challenging when patients become more active and ask questions that the provider might not always have an answer to [31]. This highlights that while promoting patient activation may be beneficial for the patient, it may also be important to consider how the provider responds to a suddenly more active and engaged patient and whether the provider needs support in adapting to this change. Another potential explanation for why providers seemingly did not report a difference in the group level of activation could be that the mean difference between the groups' CHAI-MH scores was too small for the providers to distinguish.

Although our intervention was a digital SDM tool, we did not include a specific SDM outcome measurement. This is due to our conceptual definition of SDM, defining SDM as a process rather than an outcome, and the limitation of relevant SDM measurements. While specific SDM measurements have been developed, many of these measurements are focused on a concrete decision (eg, my provider and I chose a treatment option together) rather than on the process of SDM. Challenges in measuring SDM have previously been identified, and measurements to evaluate the SDM process with adequate psychometric properties are needed [32].

On the basis of the explorative midintervention assessment, the difference in patient activation between the groups occurred between the mid- and postintervention assessment. Each group had a similar increase in CHAI-MH score from baseline to the midintervention assessment, with no between-group differences. However, only the intervention group continued to increase their CHAI-MH scores from during the intervention to after the intervention, resulting in a statistically significant difference between the 2 groups. This may indicate that the effect of a digital SDM tool may not occur quickly but instead requires

time to develop an effect. What seems contradictory is that data on app use indicate that approximately half of the participants stopped using the app after 1 month. These explorative findings suggest that participants who stopped using the app before the end of the intervention may still have benefited from it.

Our intervention group encountered both barriers and difficulties in acquiring and using an app in combination with their treatment. First, 10 participants were invited to use the app but never open it. Although we did not explore the reasons for this, there were reports of technical limitations (eg, the phone system did not support the app). The study was also affected by the fire accident at OVHcloud. Around half of the users who had used the app before the accident did not log in after the system became available again, while the mean use of the app also decreased. The magnitude of such accident is rare but does highlight a vulnerability to digital systems while also highlighting a challenge in re-engaging participants after a "pause" from an intervention. It also questions whether the effect of the digital SDM tool could have been greater if these limitations had been avoided.

Strengths and Limitations

This study had several strengths. First, we included and assessed both patients and their providers to acknowledge the importance of both in the process of SDM. Second, the study had a large sample size with a low level of lost-to-follow-up on our main outcome. Third, the study had a pragmatic nature, in which the use of the system would be similar to how it would be used in practice outside of the trial. Therefore, the results should be generalizable to other similar services.

However, this pragmatic approach is limited in terms of support for participants. Participants who encountered an issue with the app were instructed to ask their provider for assistance who were then able to consult with a blinded student assistant or an IT supporter. This placed a large responsibility on the provider. If the provider did not resolve the situation or contact support, the patient could be prone to stop using the app. A recent study highlighted that with the rapid development and use of digital tools in mental health care, educational efforts are needed to strengthen the clinician's knowledge and skills regarding these tools [33]. Future trials investigating a digital system are encouraged to carefully consider how participants (patients and providers) are supported in the case of issues or barriers.

During our recruitment, we randomized patients to either the control or intervention group, meaning that providers were able to have patients in both groups. This creates a risk for a contamination effect, as providers were able to use elements from the intervention with patients in the control group. One way to address this would have been to randomize at the clinician or clinic level to avoid providers having participants in each group. Doing so would potentially have made it more difficult to recruit participants unless they could have been to assign patients to a waitlist. However, this was not possible in this trial.

This recruitment process may challenge the generalizability of the study. The vast majority of participants were recruited by providers to inform patients about the study. Although providers were strongly encouraged to ask all of their patients about the research project, providers were able to, on their own, select which patients to inform about the study. Providers may have been more prone to ask patients they assume would use a smartphone app or patients whom the provider believed were able to participate in such a trial. This recruitment process may have affected the distribution of the study participants' characteristics, such as gender, diagnosis, or use of antipsychotic medication. For example, the level of functioning of the included participants was significantly higher in our sample than in a sample from a previous OPUS project [34]. Furthermore, this selection by the provider may show a lack of SDM between the patient and provider in which the provider decides whether to inform the patient about the research project, thereby not giving the patient a say when making the decision about participating in the study.

During the trial, researchers routinely made providers aware of the project by being physically present at the clinic. However, the COVID-19 pandemic added another challenge, in which it was not possible to be as present at the clinic as usual. However, with the pandemic, health care and many other areas saw an increased use of digital systems and how quickly we can adopt these systems into practice. eHealth provides an approach to care when in-person services are troublesome [33]. Moving forward, stakeholders and practitioners are encouraged to adopt e–mental health care tools to offer a more blended care plan [35].

Although the level of lost-to-follow-up on our primary outcome was low (<10%), we observed that the intervention group had a higher number at the postintervention time point than the control group, which may have biased our results. In addition, the higher number of participants lost to follow-up in the intervention group may be caused by boredom or dissatisfaction with the tool. However, disengagement can also be interpreted as a potentially harmful outcome of using a digital SDM tool. However, the absolute numbers of lost-to-follow-up were relatively low, and the reasons in the intervention group were mostly due to ending OPUS treatment prematurely. Another potential bias in the trial was the fact that a majority of outcomes (including the primary outcome) were self-reported, and as participants were not blinded, this could introduce a bias by overestimating the true effect size.

Conclusions

The Momentum Trial had a significant beneficial effect on the primary outcome, patient activation, at the postintervention time point (mean difference was 4.39 point favoring the intervention group with 95% CI 0.99-7.79; Cohen *d*=0.33; *P*=.01). The effect size was smaller than the 0.42 effect size that we had anticipated and sampled for. The intervention was also effective in improving secondary outcomes: confidence in communicating with one's provider and feeling prepared when making treatment decisions. Despite our hypothesis, the Momentum Trial had no effect on hope, treatment satisfaction, working alliance, or clinical outcomes.

The Momentum Trial strengthens the existing evidence by demonstrating that digital SDM interventions can be effective in supporting patients to feel active and engaged in their

treatment. This intervention had important limitations that should be considered in future trials.

Acknowledgments

The authors would like to thank all the OPUS patients and providers for their sustained interest, support, and participation in the study; TrygFonden for making the trial possible; and Odense Patient Data Explorative Network, Odense University Hospital, Odense, Denmark for assisting with randomization and statistics.

This study was funded by TrygFonden in Denmark under ID number 115441. TrygFonden had no role in the conduct of the study; collection; management; analysis and interpretation of data; review, approval, or submission of the manuscript.

Data Availability

The data generated from the trial are available from the corresponding author upon reasonable request.

Authors' Contributions

TV conducted the trial and analyzed the data under the supervision of CH. TV, LK, and CH wrote the first draft of the manuscript. TV, LK, CH, MN, SFA, EST, and LP revised and optimized the manuscript. LK has along with MN, EST, LP, and SFA taken the initiative to conduct the trial, with LK as the trial sponsor. TV, LK, SFA, LP, EST, and MN contributed to the planning of the study. All authors contributed to and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1256 KB - jmir_v24i10e40292_app1.pdf]

References

1. Salyers MP, Zisman-Ilani Y. Shared decision-making and self-directed care. In: Goldman HH, Frank RG, Morrissey JP, editors. The Palgrave Handbook of American Mental Health Policy. Cham, Switzerland: Springer; 2020:197-228.
2. Hamann J, Holzhüter F, Blakaj S, Becher S, Haller B, Landgrebe M, et al. Implementing shared decision-making on acute psychiatric wards: a cluster-randomized trial with inpatients suffering from schizophrenia (SDM-PLUS). Epidemiol Psychiatr Sci 2020 Jun 16;29:e137 [FREE Full text] [doi: 10.1017/S2045796020000505] [Medline: 32539907]
3. Zisman-Ilani Y, Roe D, Elwyn G, Kupermintz H, Patya N, Peleg I, et al. Shared decision making for psychiatric rehabilitation services before discharge from psychiatric hospitals. Health Commun 2019 May;34(6):631-637. [doi: 10.1080/10410236.2018.1431018] [Medline: 29393685]
4. Fiorillo A, Barlati S, Bellomo A, Corrivetti G, Nicolò G, Sampogna G, et al. The role of shared decision-making in improving adherence to pharmacological treatments in patients with schizophrenia: a clinical review. Ann Gen Psychiatry 2020 Aug 5;19:43 [FREE Full text] [doi: 10.1186/s12991-020-00293-4] [Medline: 32774442]
5. Waldemar AK, Esbensen BA, Korsbek L, Petersen L, Arnfred S. Recovery-oriented practice: participant observations of the interactions between patients and health professionals in mental health inpatient settings. Int J Ment Health Nurs 2019 Feb;28(1):318-329. [doi: 10.1111/inm.12537] [Medline: 30151987]
6. Hamann J, Heres S. Adapting shared decision making for individuals with severe mental illness. Psychiatr Serv 2014 Dec 01;65(12):1483-1486. [doi: 10.1176/appi.ps.201400307] [Medline: 25756970]
7. Lepping P, Stanly T, Turner J. Systematic review on the prevalence of lack of capacity in medical and psychiatric settings. Clin Med (Lond) 2015 Aug;15(4):337-343 [FREE Full text] [doi: 10.7861/clinmedicine.15-4-337] [Medline: 26407382]
8. Drake RE, Deegan PE, Rapp C. The promise of shared decision making in mental health. Psychiatr Rehabil J 2010;34(1):7-13. [doi: 10.2975/34.1.2010.7.13] [Medline: 20615839]
9. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
10. 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]
11. Vitger T, Korsbek L, Austin SF, Petersen L, Nordentoft M, Hjorthøj C. Digital shared decision-making interventions in mental healthcare: a systematic review and meta-analysis. Front Psychiatry 2021 Sep 6;12:691251 [FREE Full text] [doi: 10.3389/fpsyt.2021.691251] [Medline: 34552514]

12. Vitger T, Austin SF, Petersen L, Tønder ES, Nordentoft M, Korsbek L. The Momentum trial: the efficacy of using a smartphone application to promote patient activation and support shared decision making in people with a diagnosis of schizophrenia in outpatient treatment settings: a randomized controlled single-blind trial. BMC Psychiatry 2019 Jun 17;19(1):185 [FREE Full text] [doi: 10.1186/s12888-019-2143-2] [Medline: 31208376]
13. Nordentoft M, Melau M, Iversen T, Petersen L, Jeppesen P, Thorup A, et al. From research to practice: how OPUS treatment was accepted and implemented throughout Denmark. Early Interv Psychiatry 2015 Apr;9(2):156-162. [doi: 10.1111/eip.12108] [Medline: 24304658]
14. Korsbek L, Tønder ES. Momentum: a smartphone application to support shared decision making for people using mental health services. Psychiatr Rehabil J 2016 Jun;39(2):167-172. [doi: 10.1037/prj0000173] [Medline: 27030907]
15. Puschner B, Neumann P, Jordan H, Slade M, Fiorillo A, Giacco D, CEDAR study group. Development and psychometric properties of a five-language multiperspective instrument to assess clinical decision making style in the treatment of people with severe mental illness (CDMS). BMC Psychiatry 2013 Feb 04;13:48 [FREE Full text] [doi: 10.1186/1471-244X-13-48] [Medline: 23379280]
16. Wolf MS, Smith SG, Pandit AU, Condon DM, Curtis LM, Griffith J, et al. Development and validation of the consumer health activation index. Med Decis Making 2018 Apr;38(3):334-343 [FREE Full text] [doi: 10.1177/0272989X17753392] [Medline: 29436308]
17. Snyder CR, Sympson SC, Ybasco FC, Borders TF, Babyak MA, Higgins RL. Development and validation of the State Hope Scale. J Pers Soc Psychol 1996 Feb;70(2):321-335. [doi: 10.1037//0022-3514.70.2.321] [Medline: 8636885]
18. Schwarzer R, Jerusalem M. General Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs. Windsor, UK: Nfer-Nelson; 1995:35-37.
19. ten Klooster PM, Oostveen JC, Zandbelt LC, Taal E, Drossaert CH, Harmsen EJ, et al. Further validation of the 5-item perceived efficacy in patient-physician interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns 2012 Apr;87(1):125-130. [doi: 10.1016/j.pec.2011.07.017] [Medline: 21889864]
20. Busseri MA, Tyler JD. Interchangeability of the working alliance inventory and working alliance inventory, short form. Psychol Assess 2003 Jun;15(2):193-197. [doi: 10.1037/1040-3590.15.2.193] [Medline: 12847779]
21. Bennett C, Graham ID, Kristjansson E, Kearing SA, Clay KF, O'Connor AM. Validation of a preparation for decision making scale. Patient Educ Couns 2010 Jan;78(1):130-133. [doi: 10.1016/j.pec.2009.05.012] [Medline: 19560303]
22. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
23. Arndt S, Andreasen NC, Flaum M, Miller D, Nopoulos P. A longitudinal study of symptom dimensions in schizophrenia. Prediction and patterns of change. Arch Gen Psychiatry 1995 May;52(5):352-360. [doi: 10.1001/archpsyc.1995.03950170026004] [Medline: 7726715]
24. Pedersen G, Karterud S. The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. Compr Psychiatry 2012 Apr;53(3):292-298. [doi: 10.1016/j.comppsych.2011.04.007] [Medline: 21632038]
25. Nasrallah H, Morosini P, Gagnon DD. Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia. Psychiatry Res 2008 Nov 30;161(2):213-224. [doi: 10.1016/j.psychres.2007.11.012] [Medline: 18848731]
26. Tait L, Birchwood M, Trower P. A new scale (SES) to measure engagement with community mental health services. J Ment Health 2002;11(2):191-198. [doi: 10.1080/09638230020023570-2] [Medline: 21208145]
27. User experience - App Rating Questionnaire. R-outcomes. 2016. URL: https://r-outcomes.com/measures/ [accessed 2018-01-01]
28. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
29. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
30. van der Krieke L, Emerencia AC, Boonstra N, Wunderink L, de Jonge P, Sytema S. A web-based tool to support shared decision making for people with a psychotic disorder: randomized controlled trial and process evaluation. J Med Internet Res 2013 Oct 07;15(10):e216 [FREE Full text] [doi: 10.2196/jmir.2851] [Medline: 24100091]
31. Alegría M, Polo A, Gao S, Santana L, Rothstein D, Jimenez A, et al. Evaluation of a patient activation and empowerment intervention in mental health care. Med Care 2008 Mar;46(3):247-256 [FREE Full text] [doi: 10.1097/MLR.0b013e318158af52] [Medline: 18388839]
32. Barr PJ, Elwyn G. Measurement challenges in shared decision making: putting the 'patient' in patient-reported measures. Health Expect 2016 Oct;19(5):993-1001 [FREE Full text] [doi: 10.1111/hex.12380] [Medline: 26111552]
33. 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]

34. Albert N, Melau M, Jensen H, Emborg C, Jepsen JR, Fagerlund B, et al. Five years of specialised early intervention versus two years of specialised early intervention followed by three years of standard treatment for patients with a first episode psychosis: randomised, superiority, parallel group trial in Denmark (OPUS II). BMJ 2017 Jan 12;356:i6681 [FREE Full text] [doi: 10.1136/bmj.i6681] [Medline: 28082379]
35. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: the 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]

Abbreviations

CDMS: Clinical Decision Making Style
CHAI-MH: Consumer Health Activation Index for mental health
CONSORT: Consolidated Standards of Reporting Trials
CSQ: Client Satisfaction Questionnaire
GAF: Global Assessment of Functioning
GSE: General Self-Efficacy Scale
ITT: intention to treat
OPEN: Odense Patient Data Explorative Network
PEPPI: Perceived Efficacy in Patient-Physician Interactions
PrepDM: Preparation for Decision-Making
PSP: Personal and Social Performance Scale
SANS: Scale for the Assessment of Negative Symptoms
SAPS: Scale for the Assessment of Positive Symptoms
SES: Service Engagement Scale
SDM: shared decision-making
TAU: treatment as usual
WAI-S: Working Alliance Inventory–short form

Edited by R Kukafka; submitted 16.06.22; peer-reviewed by N Chalghaf, RJ Drake, S Desselle; comments to author 06.08.22; revised version received 12.08.22; accepted 15.09.22; published 26.10.22.

Please cite as:
Vitger T, Hjorthøj C, Austin SF, Petersen L, Tønder ES, Nordentoft M, Korsbek L
A Smartphone App to Promote Patient Activation and Support Shared Decision-making in People With a Diagnosis of Schizophrenia in Outpatient Treatment Settings (Momentum Trial): Randomized Controlled Assessor-Blinded Trial
J Med Internet Res 2022;24(10):e40292
URL: https://www.jmir.org/2022/10/e40292
doi:10.2196/40292
PMID:36287604

Corrigenda and Addenda

Correction: A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study

Géraldine Falgarone[1,2], MD, PhD; Guilhem Bousquet[1,3], MD, PhD; Arnaud Wilmet[4], MD; Albert Brizio[4], MD; Valérie Faure[4], MD; Celestin Guillouet[4], MD; Franck Baudino[4], MD; Isabelle Roque[5], MD; Samuel Mayol[6*], PhD; Frederic Pamoukdjian[1,7*], MD, PhD

[1]UMR_S942 MASCOT, INSERM, Université Sorbonne Paris Nord, Bobigny, France
[2]Unité de Médecine Ambulatoire, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[3]Service de Cancérologie, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[4]H4D (Health for Development), Paris, France
[5]Université Paris Cité, Paris, France
[6]Institut Universitaire de Technologie, Université Sorbonne Paris Nord, St-Denis, France
[7]Service de Médecine Gériatrique, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[*]these authors contributed equally

Corresponding Author:
Géraldine Falgarone, MD, PhD
Unité de Médecine Ambulatoire
Hôpital Avicenne
Assistance Publique–Hôpitaux de Paris
125 Rue de Stalingrad
Bobigny, 93009
France
Phone: 33 148026870
Fax: 33 148955258
Email: g.falgarone@aphp.fr

Related Article:

Correction of: https://www.jmir.org/2022/5/e33507

(J Med Internet Res 2022;24(10):e43220) doi:10.2196/43220

In "A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study" (J Med Internet Res 2022;24(5):e33507), the authors noted one error.

In the originally published article, author Frederic Pamoukdjian's name incorrectly appeared as:

Frederic Pamoukjian

It has now been corrected to:

Frederic Pamoukdjian

The correction will appear in the online version of the paper on the JMIR Publications website on October 13, 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 04.10.22; this is a non–peer-reviewed article;accepted 04.10.22; published 13.10.22.

Please cite as:
Falgarone G, Bousquet G, Wilmet A, Brizio A, Faure V, Guillouet C, Baudino F, Roque I, Mayol S, Pamoukdjian F
Correction: A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study
J Med Internet Res 2022;24(10):e43220
URL: https://www.jmir.org/2022/10/e43220
doi:10.2196/43220
PMID:36229035

Corrigenda and Addenda

Correction: Racial Bias Beliefs Related to COVID-19 Among Asian Americans, Native Hawaiians, and Pacific Islanders: Findings From the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS)

Van Ta Park[1,2,3*], MPH, PhD; Janice Y Tsoh[2,3,4*], PhD; Marcelle Dougan[5], MEng, MPH, SCD; Bora Nam[1], MSN, PhD; Marian Tzuang[1], MSW, PhD; Linda G Park[1], NP, PhD, FAAN; Quyen N Vuong[6], ASW, MSW, MBA; Joon Bang[7], BA; Oanh L Meyer[8], MAS, PhD

[1]Department of Community Health Systems, School of Nursing, University of California San Francisco, San Francisco, CA, United States
[2]Asian American Research Center on Health (ARCH), University of California San Francisco, San Francisco, CA, United States
[3]Multiethnic Health Equity Research Center, University of California San Francisco, San Francisco, CA, United States
[4]Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, CA, United States
[5]Department of Public Health and Recreation, San Jose State University, San Jose, CA, United States
[6]International Children Assistance Network, Milpitas, CA, United States
[7]Iona Senior Services, Washington DC, DC, United States
[8]Department of Neurology, School of Medicine, University of California Davis, Sacramento, CA, United States
[*]these authors contributed equally

Corresponding Author:
Van Ta Park, MPH, PhD
Department of Community Health Systems
School of Nursing
University of California San Francisco
2 Koret Way
San Francisco, CA, 94143
United States
Phone: 1 415 514 3318
Email: van.park@ucsf.edu

Related Article:

Correction of: https://www.jmir.org/2022/8/e38443

(J Med Internet Res 2022;24(10):e42716) doi:10.2196/42716

In "Racial Bias Beliefs Related to COVID-19 Among Asian Americans, Native Hawaiians, and Pacific Islanders: Findings From the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS)" (J Med Internet Res 2022;24(8):e38443), two errors were noted.

In the originally published article, author Linda G Park was inadvertently left out of the authorship, and the order of authors was listed as follows:

> *Van Ta Park, Janice Y Tsoh, Marcelle Dougan, Bora Nam, Marian Tzuang, Quyen N Vuong, Joon Bang, Oanh L Meyer.*

In the corrected article, author Linda G Park is listed as the sixth author, and the order of authors has been updated as follows:

> *Van Ta Park, Janice Y Tsoh, Marcelle Dougan, Bora Nam, Marian Tzuang, Linda G Park, Quyen N Vuong, Joon Bang, Oanh L Meyer.*

In the originally published article, the phone number of the Corresponding Author was incorrect. The phone number has been corrected to:

> *1 415 514 3318*

The correction will appear in the online version of the paper on the JMIR Publications website on October 31, 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 14.09.22; this is a non–peer-reviewed article;accepted 15.09.22; published 31.10.22.

Please cite as:
Park VT, Tsoh JY, Dougan M, Nam B, Tzuang M, Park LG, Vuong QN, Bang J, Meyer OL
Correction: Racial Bias Beliefs Related to COVID-19 Among Asian Americans, Native Hawaiians, and Pacific Islanders: Findings From the COVID-19 Effects on the Mental and Physical Health of Asian Americans and Pacific Islanders Survey Study (COMPASS)
J Med Internet Res 2022;24(10):e42716
URL: https://www.jmir.org/2022/10/e42716
doi:10.2196/42716
PMID:36315935

Original Paper

Emotions and Topics Expressed on Twitter During the COVID-19 Pandemic in the United Kingdom: Comparative Geolocation and Text Mining Analysis

Hassan Alhuzali[1*], PhD; Tianlin Zhang[2*], MSc; Sophia Ananiadou[2,3], PhD

[1]College of Computers and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

[2]Department of Computer Science, National Centre for Text Mining, The University of Manchester, Manchester, United Kingdom

[3]The Alan Turing Institute, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sophia Ananiadou, PhD
Department of Computer Science
National Centre for Text Mining
The University of Manchester
131 Princess Street
Manchester, M1 7DN
United Kingdom
Phone: 44 161 306 3092
Email: sophia.ananiadou@manchester.ac.uk

Abstract

Background: In recent years, the COVID-19 pandemic has brought great changes to public health, society, and the economy. Social media provide a platform for people to discuss health concerns, living conditions, and policies during the epidemic, allowing policymakers to use this content to analyze the public emotions and attitudes for decision-making.

Objective: The aim of this study was to use deep learning–based methods to understand public emotions on topics related to the COVID-19 pandemic in the United Kingdom through a comparative geolocation and text mining analysis on Twitter.

Methods: Over 500,000 tweets related to COVID-19 from 48 different cities in the United Kingdom were extracted, with the data covering the period of the last 2 years (from February 2020 to November 2021). We leveraged three advanced deep learning–based models for topic modeling to geospatially analyze the sentiment, emotion, and topics of tweets in the United Kingdom: SenticNet 6 for sentiment analysis, SpanEmo for emotion recognition, and combined topic modeling (CTM).

Results: We observed a significant change in the number of tweets as the epidemiological situation and vaccination situation shifted over the 2 years. There was a sharp increase in the number of tweets from January 2020 to February 2020 due to the outbreak of COVID-19 in the United Kingdom. Then, the number of tweets gradually declined as of February 2020. Moreover, with identification of the COVID-19 Omicron variant in the United Kingdom in November 2021, the number of tweets grew again. Our findings reveal people's attitudes and emotions toward topics related to COVID-19. For sentiment, approximately 60% of tweets were positive, 20% were neutral, and 20% were negative. For emotion, people tended to express highly positive emotions in the beginning of 2020, while expressing highly negative emotions over time toward the end of 2021. The topics also changed during the pandemic.

Conclusions: Through large-scale text mining of Twitter, our study found meaningful differences in public emotions and topics regarding the COVID-19 pandemic among different UK cities. Furthermore, efficient location-based and time-based comparative analysis can be used to track people's thoughts and feelings, and to understand their behaviors. Based on our analysis, positive attitudes were common during the pandemic; optimism and anticipation were the dominant emotions. With the outbreak and epidemiological change, the government developed control measures and vaccination policies, and the topics also shifted over time. Overall, the proportion and expressions of emojis, sentiments, emotions, and topics varied geographically and temporally. Therefore, our approach of exploring public emotions and topics on the pandemic from Twitter can potentially lead to informing how public policies are received in a particular geographical area.

(J Med Internet Res 2022;24(10):e40323) doi:10.2196/40323

KEYWORDS

Twitter; COVID-19; geolocation; emotion detection; sentiment analysis; topic modeling; social media; natural language processing; deep learning

Introduction

The crisis of the COVID-19 pandemic has influenced the whole world on an enormous scale, causing most countries to deal with an unprecedented situation. The societal consequences due to lockdowns were tremendous on all levels. The pandemic caused most countries to impose various stages of restrictions on moving, traveling, and gathering to contain the outbreak of infection. Such restrictions changed how people used to work, socialize, shop, travel, etc, leading to various behavioral and societal changes to deal with the situation (eg, working from home, fear of social interaction, isolation, loneliness). Because of this unprecedented societal change, it was important for policymakers to understand people's state of mind to help institutions, governments, and individuals navigate through the pandemic [1-4].

Traditionally, policymakers used questionnaires to capture public opinion toward major events, but there are disadvantages limiting the effectiveness of such methods of evidence gathering due to bias caused by spatiotemporal granularity and sample sizes. Recently, social media have become an important vehicle of gathering information and evidence about public opinion. Twitter is a popular social media platform with more than 19 million users in the United Kingdom [5], where there are many discussions and opinions about topics related to COVID-19. Previous studies show that Twitter can yield important public health information and has broad applicability for public health research, including medical well-being and tracking infectious disease outbreaks [6,7]. Therefore, to address the evidence gap from traditional surveys, Twitter data can be used to supplement data gathering, and to understand public opinion on pandemics [8,9] and reactions to the COVID-19 outbreak [10].

There is a growing body of research that has recently focused on the COVID-19 pandemic with respect to different attributes, including sentiment, emotions, and topics [11-16]. Kleinberg et al [11] built the COVID-19 Real World Worry Dataset, which is based on a direct survey written by 2500 participants who reported their feelings while writing. Gupta et al [13] created another COVID-19 data set from Twitter by using a set of keywords related to the pandemic, as well as analyzing sentiment and topics as additional attributes to emotion. For instance, there are some analyses of COVID-19 vaccine–related discussions on Twitter or Reddit based on sentiment analysis and topic modeling in different countries, including the United States [17-19], Canada [20], the United Kingdom [18], Saudi Arabia [21], and Australia [22].

Sentiment represents the attitude and feelings expressed by people. Sentiment analysis determines and interprets whether online posts collected from social media are positive, neutral, or negative, and helps to gain better insight into public perceptions and attitudes. Sentiment analysis can also help to understand how information spreads on social media: a tweet with positive/negative sentiment generates another tweet with the same or opposing sentiment [23]. Sentiment analysis has been used for many practical applications, including financial analysis, politics, health prediction, and health care service improvement [24]. For instance, by analyzing public messages, sentiment analysis can be used by health practitioners to understand potential obstacles to population-based intervention approaches such as COVID-19 vaccination. In addition, analyzing patients' online reviews of different treatments can improve patient satisfaction [25].

Emotion detection from social media plays an important role in monitoring health and well-being [26]. Clinicians and health professionals also benefit from emotion analysis to understand public emotions and public health changes in perception of an intervention (ie, vaccine). Emotion detection systems have been used for alerting public health practitioners, for monitoring mental health patients [27], suicide prevention [28], and adverse drug reactions [29]. Some works utilized emotion-based features to specifically detect adverse drug reactions reported by users on social media, which can guide health professionals and pharmaceutical companies in making medications safer and advocating for patient safety [30-32]. Moreover, the idea of emotional contagion can further play a crucial role in either improving the overall well-being of users or preventing them from developing mental health problems. Kramer et al [33] stated that emotions can be transferred to others through *emotional contagion*. Emotional contagion makes people experience similar emotions, even if they are not aware of their emotional changes. On the one hand, other works found a strong link between people's mental health problems (ie, depression and anxiety) and the outbreak of COVID-19 due to the intense exposure to negative content on social media [34,35]. On the other hand, one can also expose people to positive or desired emotions (eg, calm, joy, optimism, and rest) to improve their overall well-being [33].

Besides sentiment analysis and emotion detection, topic modeling is an important text analysis technology by grouping texts into different themes. Most models can find hidden topics without supervision, and therefore do not require training on specific data with predefined topics, which makes this approach suitable for analyzing social media data to determine what people are talking about on these platforms. Topic modeling has been used for many health applications during the COVID-19 pandemic [36], such as monitoring people's concerns, predicting COVID-19 cases, and analyzing government responses. Topic modeling has played a crucial role in health information surveillance and public opinion monitoring [37].

Given the growing interest of research in understanding people's opinions and emotions regarding the pandemic [37], the objective of this study was to use deep learning–based methods to understand public emotions on topics related to the COVID-19 pandemic in the United Kingdom through a comparative geolocation and text mining analysis on Twitter.

Specifically, we utilized three advanced deep learning–based methods (ie, SenticNet [38], SpanEmo [39], and combined topic modeling [CTM] [40]), and then performed our analysis on a data set collected from Twitter to explore people's sentiment, emotions, and topics toward COVID-19. We further included analyses of these attributes focused on understanding the impact of the pandemic over time. The overall goal of this study was to automatically capture the impact COVID-19 had on the UK population using emotion detection, sentiment analysis, and topic modeling.

Methods

Data Source

To develop our corpus, we used the Twitter application programming interface by collecting data via the use of several bounding boxes over multiple cities in the United Kingdom. We further used a list of keywords that are of relevance to the pandemic (eg, coronavirus, sars19, covid19, and NHS [National Health Service]). The data covered the period of the last 2 years (ie, 2020 and 2021). To acquire location labels on the data, we used the Python geocoding library "geopy" [41], which helps locate the coordinates of addresses (eg, Oxford Rd, Manchester M13 9PL), cities (eg, Manchester), countries (eg, United Kingdom), and landmarks (in the form of latitude and longitude coordinates) based on third-party geocoders and several other data sources. More specifically, we use "Nominatim" [42] as a third-party tool. As a result, we acquired a total of 516,427 tweets from 48 cities in this study.

The number of tweets per city and emoji is shown in Table 1 and Multimedia Appendix 1, respectively. We further highlight the 9 cities that were used for our analysis: Birmingham, Bristol, Leeds, Leicester, Liverpool, London, Manchester, Nottingham, and Sheffield. It is worth mentioning that these 9 cities are also among the top populated cities in the United Kingdom [43]. This shows that there is a link between the population size and the number of posted tweets from a given geolocation area. Multimedia Appendix 1 displays the top 50 tweets (according to percentage) associated with each individual emoji and its meaning, highlighting the usage of emojis expressing different health issues (eg, virus, face with medical mask, syringe, or vaccine) and mental health conditions (eg, hands pressed together).

Table 1. Number of tweets per city in the United Kingdom.

City	Tweets, n	Population, n
Bath	1698	105,730
Birmingham[a]	21,120	1,159,888
Blackburn	1092	121,475
Bradford	4980	368,485
Brighton	10,092	245,504
Bristol[a]	10,338	580,199
Cambridge	6894	149,155
Canterbury	2292	64,495
Carlisle	1098	74,536
Chelmsford	3894	119,468
Chester	3516	87,881
Chichester	864	31,881
Coventry	6072	388,793
Derby	3503	264,430
Durham	9414	56,920
Ealing	4914	340,341
Ely	432	20,333
Exeter	3360	127,709
Gloucester	1740	148,167
Hereford	1134	64,037
Kingston	5286	287,705
Kirklees	3156	441,290
Lancaster	876	52,935
Leeds[a]	11,628	516,298
Leicester[a]	19,818	472,897
Lichfield	792	34,686
Lincoln	4614	107,434
Liverpool[a]	15,876	589,774
London[a]	111,667	9,088,994
Luton	2658	222,043
Manchester[a]	25,260	567,334
Newcastle	9642	290,688
Northampton	3954	230,070
Norwich	4290	199,245
Nottingham[a]	11,827	320,536
Peterborough	2054	179,349
Plymouth	2736	240,297
Portsmouth	4878	248,748
Preston	3816	100,095
Redbridge	3227	310,330

City	Tweets, n	Population, n
Ripon	138	15,971
Rochdale	1415	114,511
Rotherham	198	111,158
Salford	8034	125,983
Sheffield[a]	15,582	557,039
Southampton	7806	270,333
Worcester	3492	101,816
York	5748	164,934

[a]Top nine cities used in subsequent analyses.

Methodology

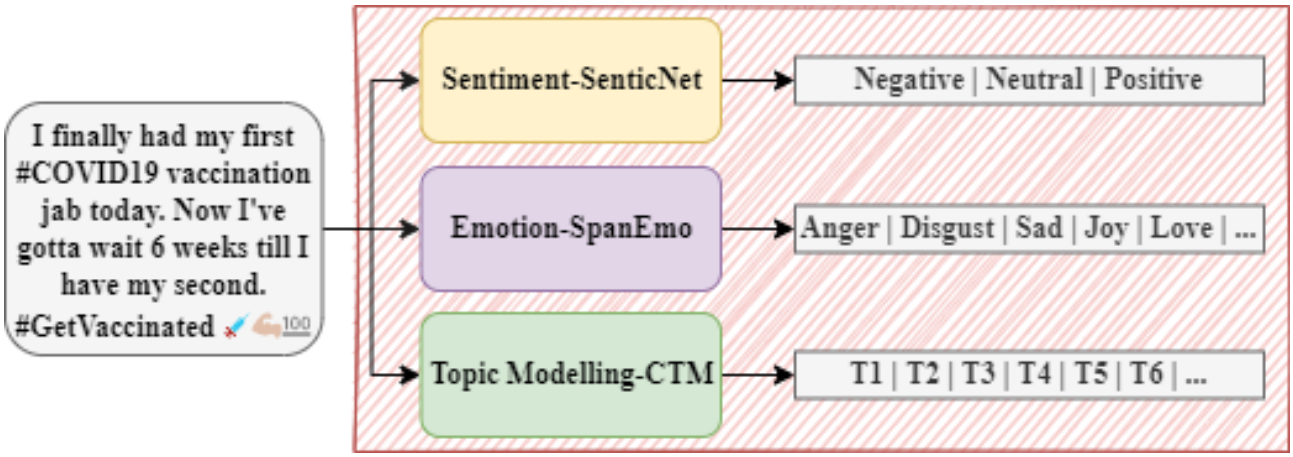
To preprocess the data, we used the "ekphrasis" tool designed for the specific characteristics of Twitter (ie, misspellings and abbreviations) [44]. The tool provides different functionalities such as tokenization, normalization, and spelling correction. We utilized the tool to tokenize the text; convert words to lowercase; and normalize user mentions, URLs, and repeated characters. Once the preprocessing step was complete, we fed the data through three models: (1) a textual emotion deep learning–based recognition model, (2) a deep learning–based sentiment model, and (3) a neural network topic model. Figure 1 depicts our pipeline, in which we provide an illustration of the three deep-learning models.

We used SenticNet 6 [38] for sentiment analysis, since this model has achieved better performance than other machine learning–based sentiment analysis methods. SenticNet 6 can provide sentiment scores (between –1 and 1) for approximately 200,000 common-sense concepts by using both symbolic models (ie, logic and semantic networks) and subsymbolic methods with deep learning architectures to encode the meanings and syntactic relations. We then added up the sentiment scores of each concept in the post and used two basic linguistic patterns (negation and adversative patterns) [45]. For example, if the patterns are not used, "The television is old but rather not expensive" could be wrongly classified although both "old" and "expensive" are negative. Finally, we calculated the sentiment polarity of each post automatically. We divided our data into five categories based on the following score range: strong negative (–1 to –0.5), weak negative (–0.5 to –0.1), neutral (–0.1 to 0.1), weak positive (0.1 to 0.5), and strong positive (0.5 to 1).

The emotion recognition model is based on our deep learning–based model "SpanEmo" [39] that is designed for multilabel emotion classification. This model is specifically trained on the SemEval-2018 multilabel emotion classification data set [46], labeled with multiple emotions classes (ie, anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust). SpanEmo focuses on both learning emotion-specific associations and integrating their correlations into the training objective. Since SpanEmo achieved strong performance for the task of multilabel emotion classification, we decided to use it to generate predictions for our data. It should be mentioned that only examples with high predictions are retained.

Last, for topic modeling, we used CTM [40]. This model incorporates contextualized document embeddings into neural topic models to produce more coherent and meaningful topics. Because the evaluation results on five publicly available data sets illustrate that the performance achieved by CTM is better than traditional latent Dirichlet allocation [47] topic models and other neural models, we employed CTM to extract the topics and their associated words from our data.

Figure 1. Overview of our pipeline. CTM: combined topic modeling.



Ethical Considerations

Since our data were collected from Twitter, we followed Twitter's terms of service and strict ethical research protocols similar to the guidelines [48], protecting the privacy and security of personal data. It should be mentioned that our study was focused on the tweet level; we do not anticipate any negative ethical impact from our analysis. However, we believe that these results provide insights into people's emotions and topics among different cities in the United Kingdom during the COVID-19 pandemic.

Results

Words Associated With Emotions

We performed different types of analyses focused on sentiment, emotion, and topic modeling of the COVID-19 online data sets. First, we analyzed emotion-words and topic-words associations where both demonstrate the relationship between words and their respective emotion label and topic. We then analyzed where the location is given, and where the impact of COVID-19 on different cities in the United Kingdom is discussed. Furthermore, an analysis of time-based features was undertaken, focusing on showing the impact of COVID-19 over time. Finally, we analyzed instances from our data that discuss the benefits of considering sentiment, emotion, and topical analysis in understanding the concerns of people during the pandemic in the United Kingdom.

Table 2 presents the top 6 words associated with each emotion and learned by SpanEmo. More detailed information on how to generate these words is provided by Alhuzali and Ananiadou [39]. There were words that are indicative of both the corresponding emotion as well as the COVID-19 pandemic. For instance, the words "death" and "spread" were highly associated with the emotion class fear, whereas words such as "vaccine" and "support" were highly associated with the emotion class anticipation. This is intuitive since some words directly express emotion (eg, angry, afraid, and glad), while other words indirectly express emotion (eg, accident, failure, and birthday). We also observed that some emotion classes shared similar words, especially those that belong to the same valence space [49]. The analysis presented in Table 2 demonstrates that it is possible to understand the impact of COVID-19 with the help of emotion analysis and the concerns of people during the pandemic.

We extracted topics using CTM. Table 3 summarizes the top 18 topics extracted as well as the top 5 associated words per topic. We noticed that there were many different topics mentioned by users, ranging from those related to COVID-19, such as epidemic control, government policies, and vaccination, to indirectly related subjects such as work, online, and social networking. For example, topic 1 (t1) contains some words about gratitude (ie, *grateful*, *thank)*, which is related to the attitude toward social support and vaccination. Topic 3 (t3) is about the discussion during the pandemic, topic 10 (t10) centers on the serious consequences of COVID-19 (*die*, *killing*), and topic 8 (t8) reveals occupational patterns.

Table 2. The top 6 words associated with each emotion class, predicted by SpanEmo.

Emotion class	Associated words
Negative emotions	
Anger	death, think, public, virus, don't, against
Disgust	deaths, virus, against, because, public, after
Fear	deaths, spread, symptoms, coronavirus, identify, self-reporting
Sadness	deaths, going, cases, hospital, other, please
Pessimism	sadly, family, friend, during, weeks, passed
Positive emotions	
Anticipation	support, vaccine, first, working, public, cases
Joy	great, thank, support, happy, amazing, staysafe
Trust	trust, thank, protect, important, community, everyone
Love	happy, loved, share, beautiful, wonderful, amazing
Optimism	please, thank, support, working, great, spread
Surprise	shocking, surprised, amazing, public, absolutely, deaths

Table 3. Topics extracted using combination topic modeling and the top 5 associated words per topic.

Topic	Associated words
t1	thank, grateful, proud, amazing, heroes
t2	class, sign, trade, worldwide, hold
t3	discuss, blog, discussion, recovery, opportunities
t4	united, fitness, kingdom, complete, image
t5	episode, tune, film, videos, radio
t6	rear, accord, whack, discomfort, fills
t7	vaccination, vaccine, dose, drug, booster
t8	letter, homes, worker, pay, private
t9	visit, eye, tweet, click, website
t10	die, dying, true, killing, cause
t11	confirmed, total, English, wales, reports
t12	rear, accord, jeopardise, unknowingly, discomfort
t13	lies, cummings, press, leader, prime
t14	coronavirus, pandemic, outbreak, instagram, outbreak
t15	masks, wear, face, hand, covering
t16	slow, thread, implement, testandtrace, symptom
t17	couple, havent, felt, daughter, holiday
t18	stay, loved, tough, pray, healthy

Analysis of Location

Figure 2 shows the number of emojis across a sample of UK cities, where the sample consists of the top 9 cities in our data, more specifically those that had the highest number of tweets (Table 1): Bristol, Birmingham, Leicester, Leeds, Liverpool, London, Manchester, Nottingham, and Sheffield. The emoji set included the following topics: virus, face-mask, thumbs-up/-down, broken heart, and others. The proportion of emojis differed from city to city. For example, usage of the syringe, or known today as the COVID-19 vaccine emoji, was high in Liverpool; the thumbs-down emoji was high in Birmingham; and the mask emoji was highly used in London and Liverpool. These emojis are relevant to the COVID-19 pandemic, demonstrating the benefits of our data in mining and analyzing social data such as Twitter for a better understanding of the impact of the pandemic on people from different areas in the United Kingdom.

In Figure 3, we present the proportions of five sentiments (strong positive, weak positive, neutral, weak negative, and strong negative) in the top 9 cities in our data in terms of their number of tweets. We can observe that approximately 60% of tweets were positive and 20% were negative in each city. At the same time, the percentage of tweets with different sentiments differed among these cities. For example, Leeds had a relatively high proportion of strong negative tweets and Sheffield had a relatively low proportion of strong positive tweets.

In Figure 4, we present the distribution of emotion expressions across the top 9 cities in our data. It can be observed that these 9 cities shared quite similar distributions, although the proportion differed from emotion to emotion. For instance, “optimism” and “anticipation” were the most frequently expressed emotions. We also noted some mixed emotions such as joy, disgust, and anger, which are reasonable feelings to be expressed during the COVID-19 pandemic. Interestingly, the proportion of trust expressions was extremely low, which could be linked to the lack of trust in decision-makers to deal properly with the situation due to the high infection rates. It is noteworthy that the proportion of trust expressions has been found to be generally scarce on Twitter in previous work [50,51].

In addition, we also counted the proportion of 10 topics in different cities, as shown in Figure 5. Similar topics received different degrees of attention in different cities. For instance, the main topic discussed in Leicester was t2 (*trade, worldwide*), which revealed that the public is more concerned about international trade. In London, the residents talked more about t4 (*kingdom, united*) than in other cities. In addition, Sheffield’s population focused more on the death topic given the higher proportion of t10 (*die, killing*) than found in the other cities.

Figure 2. The number of emojis used across a sample of UK cities.

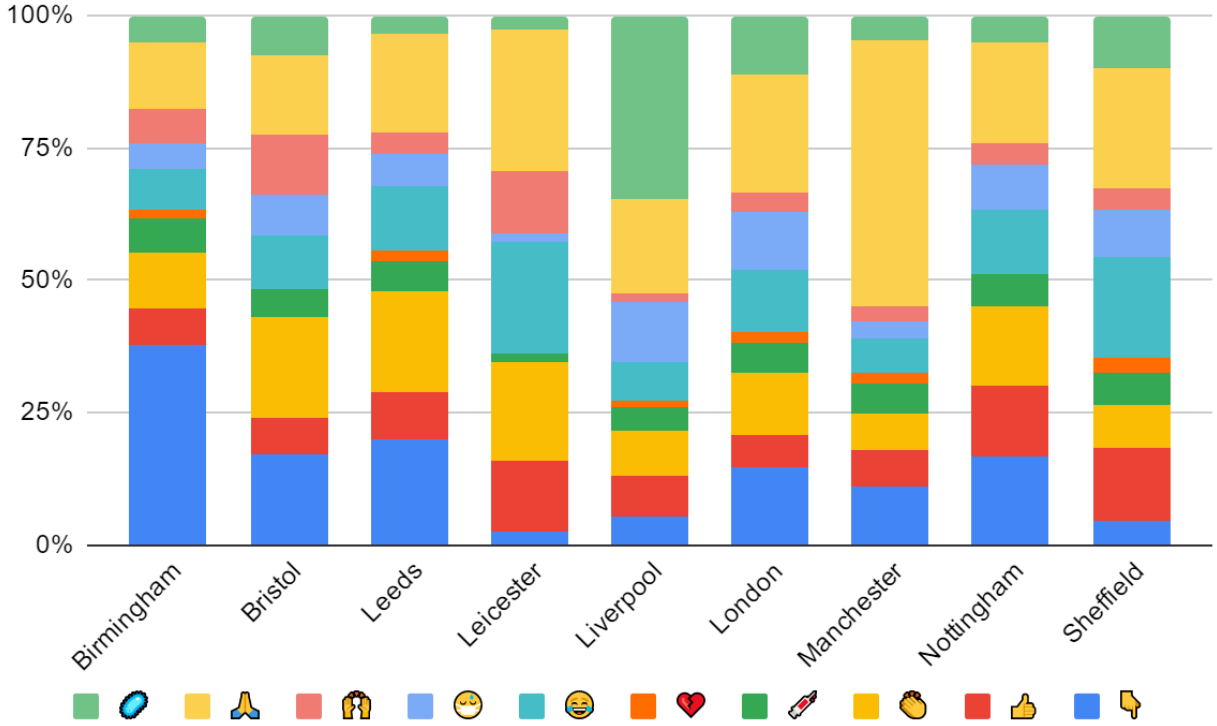


Figure 3. The distribution of sentiment expressions across a sample of UK cities.

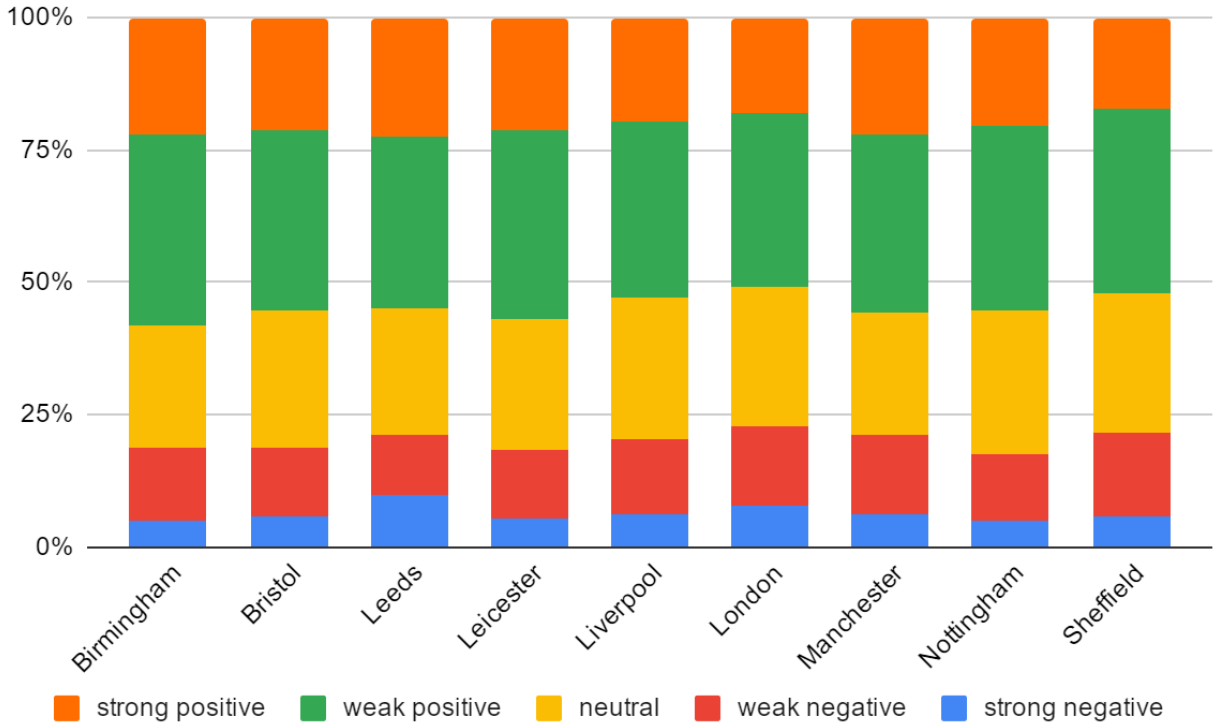


Figure 4. The distribution of emotion expressions across a sample of UK cities.

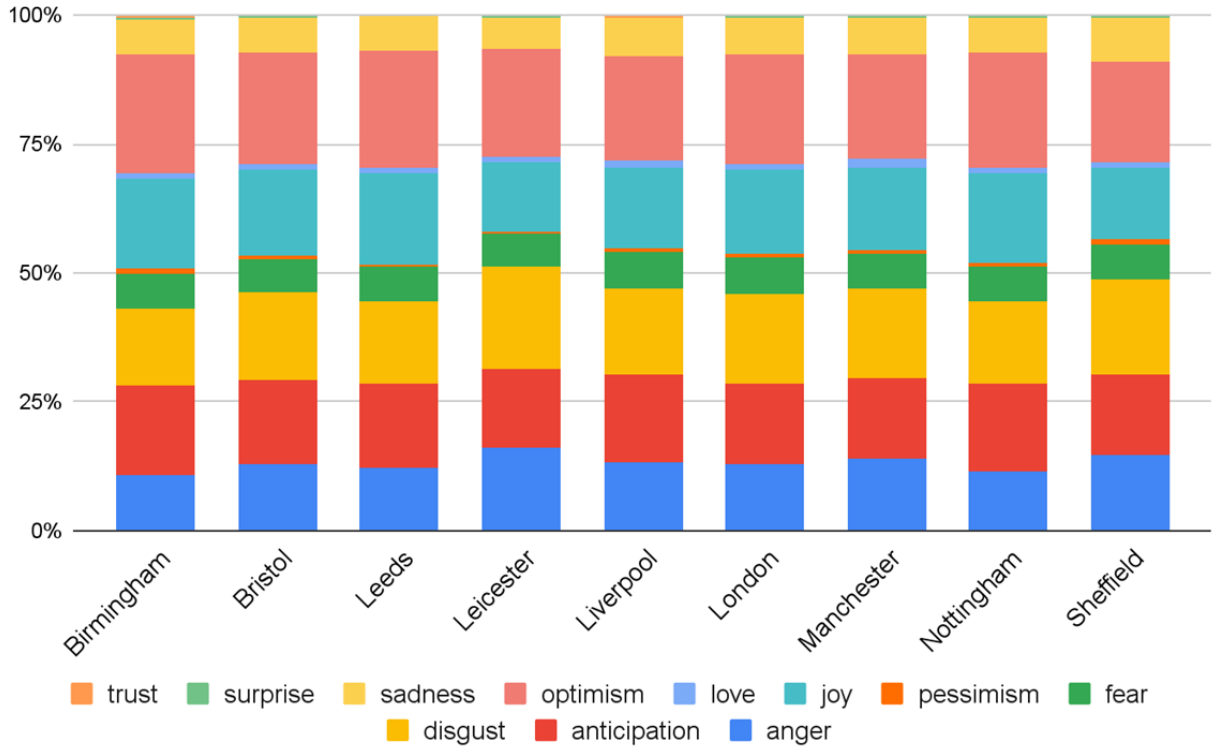
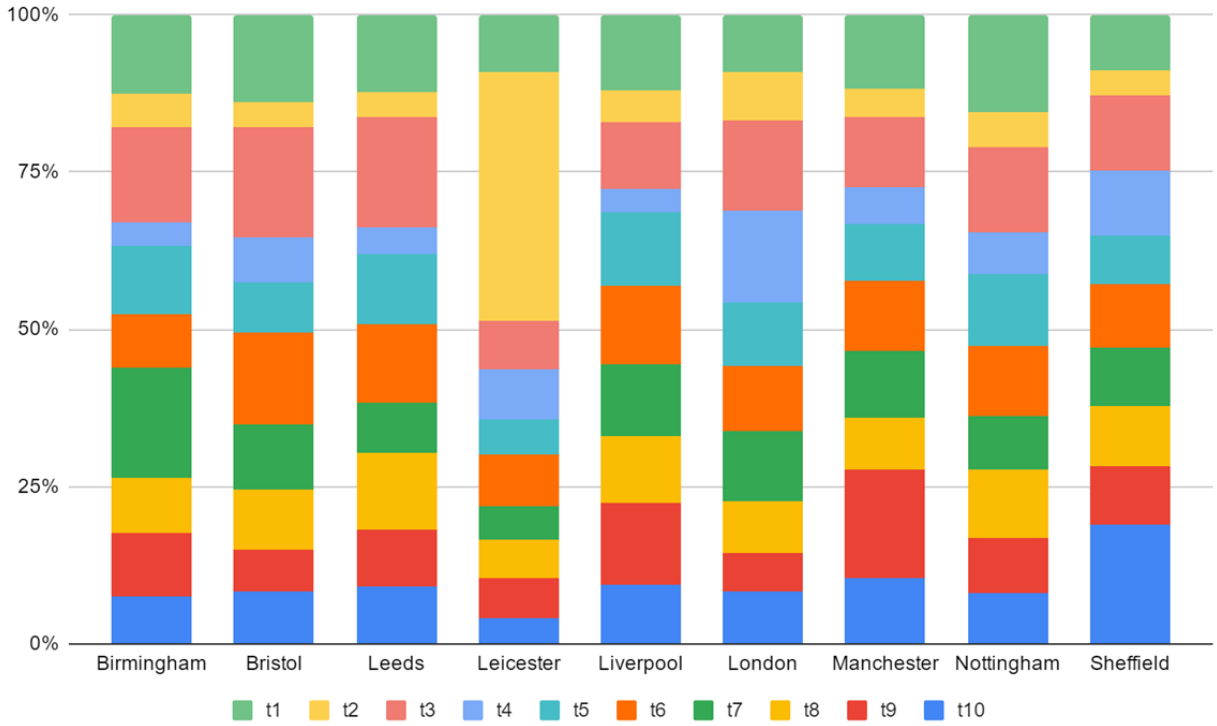


Figure 5. The distribution of topic expressions across a sample of UK cities. See Table 3 for a description of topics t1-t10.



Analysis of Time

With time, the situation of the epidemic has changed, reflecting the level of concern about the epidemic. Figure 6 displays the number of tweets related to COVID-19 from January 2020 to December 2021. We can observe a sharp increase in the number of tweets from January 2020 to February 2020 (approximately 100,000 tweets), mainly due to the outbreak of COVID-19 in the United Kingdom. There was a gradual decline in the number of tweets as of February 2020, suggesting that people became

less concerned about the epidemic. Moreover, the overall number of tweets was relatively low in 2021. With identification of the COVID-19 Omicron variant in the United Kingdom in November 2021, the number of tweets posted increased.

Figure 7 presents the emotion expressions over time, covering the 2 years (ie, 2020 and 2021). We noticed that the distribution changed with time. In the beginning of 2020, almost all emotion labels displayed high peaks of expressions, with some obviously higher than others, such as optimism. As time progressed, the number of posted tweets containing emotions decreased, but the emotion distributions had dramatically changed from being highly positive to negative. This trend progressed until reaching the end of 2021. For instance, disgust, sadness, and hopelessness were among the top expressed emotions during this period, which were reasonable emotions to be expressed during this period since the number of cases and deaths increased [52].

Figure 8 shows the change in topics (among 10 selected topics) of all tweets between February 2020 and November 2021. We can see that the change is relatively significant. In April 2020, many tweets expressed gratitude to heroes of local councils for the epidemic, given the highest frequency of messages related to t1 (*grateful*, *thanks*). In addition, due to advances in vaccine research and an increase in the number of people vaccinated, the number of tweets referring to t7 (*vaccination*) relatively increased, and reached the highest value in January 2021. Interestingly, there were many tweets related to t5 (*film*, *videos*) because of the emergence of films with special significance, such as *A Beacon of Hope: The UK Vaccine Story* and *One Year On: A pandemic poem for Londoners*. For example, someone posted “What an honour to be filmed by @BBCLondonNews reading this part of our One Year On, poem marking the anniversary of the 1st lockdown.”

Figure 6. Number of tweets related to COVID-19 from January 2020 to December 2021. Each colored line represents a specific year (ie, red represents 2020, while orange represents 2021).

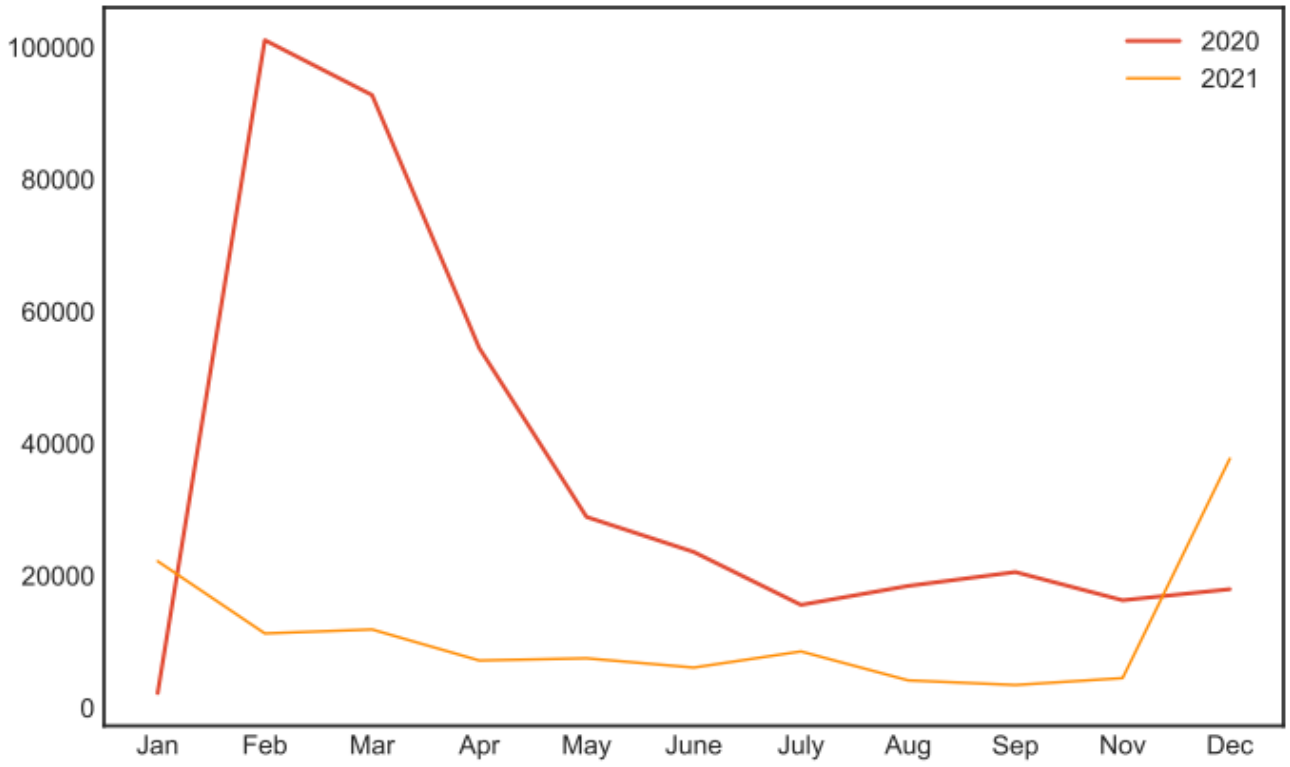


Figure 7. Number of tweets with different emotion expressions from 2020 to 2021.

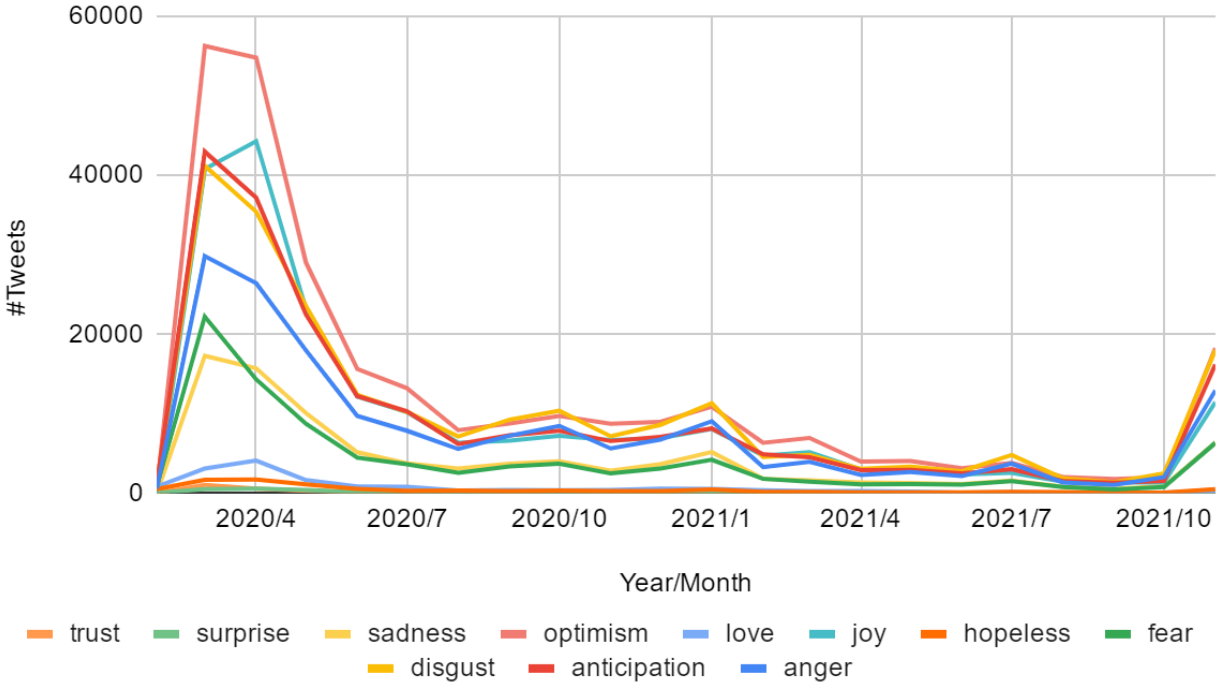
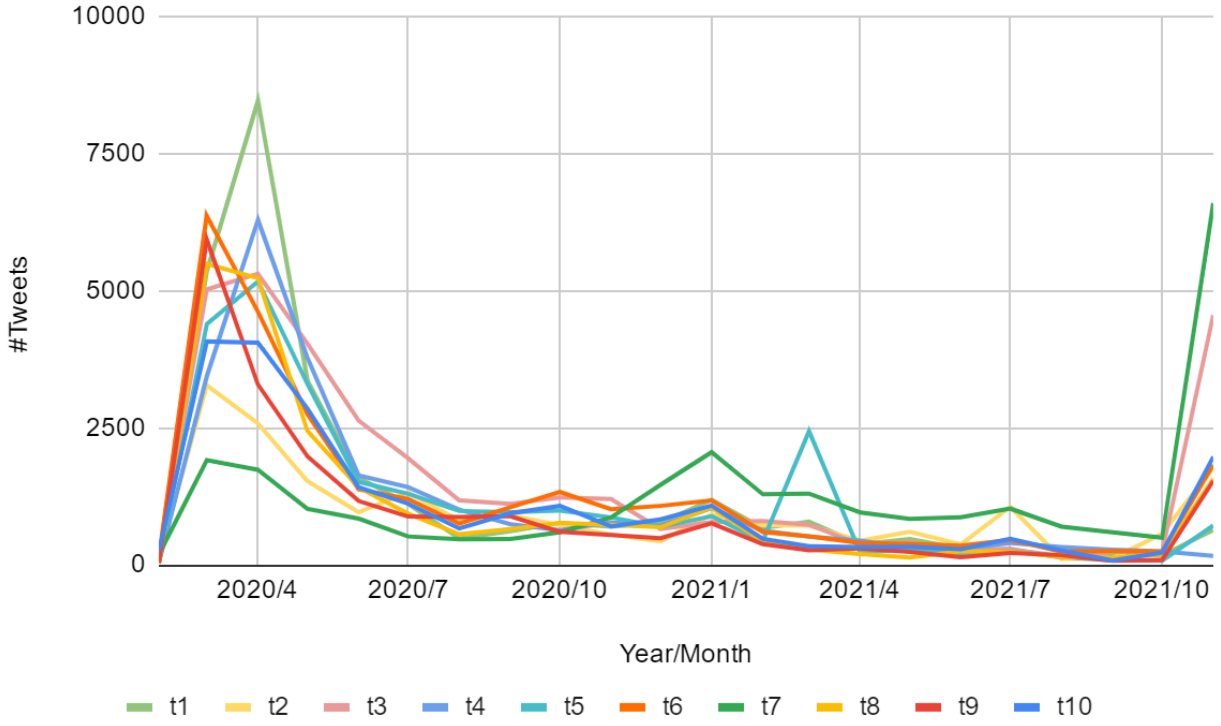


Figure 8. Number of tweets with different topic expressions from 2020 to 2021. See Table 1 for the descriptions of topics t1-t10.



Analysis of Examples

Multimedia Appendix 2 presents 9 instances from our data, each of which is linked to different attributes (ie, emotions, emojis, sentiment, and topics), demonstrating interesting findings that highlight the benefits of these attributes to the understanding of people's reactions with respect to the pandemic. Here, we describe some use cases of emojis in tweets that were commonly observed across our data. Examples 1 and 3 display the use of emojis that are related to vaccine-taking (syringe emoji) and feeling strong/protected (muscle emoji). These two examples suggest that being vaccinated can make people feel strong and protected against the COVID-19 disease. Other examples (ie, Examples 4 and 5) discuss flight

cancellation (airplane emoji), causing people to miss their already planned trips and holidays. Example 5 also discusses the potential of being able to travel again once the COVID-19 vaccine has been taken. A further example illustrates the benefits of developing volunteering programs that can assist hospitals and communities in fighting the COVID-19 crisis. Furthermore, the mask emoji was used in different ways, depending on the context (eg, lockdown for a long period).

From the perspective of sentiment, different tweets expressed different sentiments (including positive, neutral, and negative sentiments). Example 3 discusses that the second COVID-19 vaccine had been successfully administered and Example 9 praises community groups for their help and support, both of which show strong positive sentiment from the users. Example 7 expresses negative sentiment since the user could not see her relatives due to the epidemic. Some other examples (ie, Examples 1, 2, 5, 6, and 8) generally express positive attitudes during the pandemic by introducing vaccination, lockdown, or volunteers. In addition, Example 4 shows an instance that expresses mixed sentiment (positive and negative), although it was labeled by SenticNet as neutral. However, SpanEmo identified some mixed emotions, which helps to overcome the limitation of SenticNet in dealing properly with expressions having mixed sentiments or emotions.

Multimedia Appendix 2 also shows the top 3 topics for each example according to the probability calculated based on CTM. Examples 1, 3, and 5 belong to t7, which dominates the discussion of vaccination and boosters. Examples 4, 7, and 8 express the users' attitudes and moods toward the impact of COVID-19 on their lives, and thus all of these were classified as t18. Examples 3 and 6 also belong to t1 (related to gratitude) because of the appearance of "thank you." Moreover, discussion or usage of social media (t3) was expressed in some tweets (eg, Example 9).

Discussion

Principal Findings

This study explored more than 500,000 tweets related to COVID-19 between January 2020 and December 2021 in different cities of the United Kingdom, where the number of tweets increased dramatically following the outbreak. We used three deep learning–based models to analyze and combine sentiments, emotions, and topics to identify the key public concerns during the pandemic. Through our analysis, we found that emotion analysis can support understanding of people's opinions and attitudes during the COVID-19 pandemic. Meanwhile, taking geolocation information into account can reveal differences between different areas in the United Kingdom. The overall sentiment was positive over time, and optimism was the predominant emotion, suggesting that people tend to be optimistic about the situation. There were changes in the sentiments, emotions, and topics expressed on Twitter as the epidemiological situation and government policies changed (eg, vaccination, social distancing) over these 2 years, which also reflect changes in people's attitudes.

The benefits of the selected attributes for gathering evidence about people's reactions during the pandemic in the United Kingdom were also identified. These attributes include emotion, sentiment, emojis, and topic modeling. This analysis demonstrated that such attributes can help gather evidence and analyze interactions between people during the pandemic. The first attribute was emotion, which can serve as a guide in understanding people's reactions. For example, some people express concerns about COVID-19 for multiple reasons such as (1) taking a longer time to be resolved than expected, (2) cancelling or changing plans, (3) traveling restrictions, (4) wearing masks, and (5) isolation and lack of contact from family and friends. Others express some positive reactions and potential solutions for dealing with the pandemic, including family support, being inoculated with vaccines, staying at home or wearing masks, and volunteering. The second attribute was emojis, which describe the overall expression in the text, similar to topic modeling in the sense that both refer to the topics expressed in tweets. This provides another dimensionality for emojis, which have been used as a surrogate to collect emotion data [53,54]. Although this point is interesting to observe through this work, we leave it for future work to be investigated in greater depth.

Sentiment analysis is also useful to gain insight into the public opinion and perception behind certain events. By analyzing the sentiments in our data, we found that most people have had a positive attitude during the pandemic, which matches the conclusion of previous research [55], since they often post information related to good policies such as social support and vaccination to boost confidence in the fight against COVID-19. Definitely, some people still expressed worry about the outbreak and developed negative feelings due to the deaths, isolation, and lockdown policies, which affected their normal lives.

From the topics extracted, we found that there are many distinct topics people focus on, including symptoms of COVID-19, vaccination, social media, government policies, and living conditions. The changing themes of social media reveal the impact of COVID-19 on people's lives, shifting the discussion about daily life to the pandemic and policies.

In addition, the emojis used, the emotions expressed, and the topics discussed by people who are from different cities in the United Kingdom all differed because of various factors such as the environment in the city, the epidemic situation, policies, and hot spots. The findings reveal the complexity and diversity of people's perceptions toward the COVID-19 pandemic, which indicates the need to keep track of public attitudes.

Limitations

This work is based on existing natural language processing methods that were used to analyze different attributes such as emotions, sentiment, and topics. However, these existing methods may not guarantee that their predictions reflect the actual attribute. In addition, emotion and sentiment are subjective tasks, which make them difficult to model and in turn could affect our interpretation as well as our results. Moreover, since our data were collected from Twitter with the use of specific keywords, it is possible that we missed other topics in online threads and viewpoints. Related discussions

could also be taken from other social media platforms (eg, Facebook, Reddit). In this respect, our data provide a partial sample of user interactions on Twitter. The methods nevertheless are applicable to other longitudinal data and social media platforms.

Conclusion

Our main contribution is the multimethod approach that provides insights into public sentiment and emotions in UK cities during the COVID-19 pandemic. Furthermore, our methods are location- and time-based, supporting a comparative analysis to track public concerns. Our analysis demonstrated that positive attitudes were common during the pandemic; optimism and anticipation were the dominant emotions. With the outbreak and epidemiological change, the government developed control measures and vaccination policies, and the topics also shifted over time. In addition, the comparative geolocation analysis revealed differences in the emotions expressed and topics discussed by people in different cities. Overall, our study shows that analyzing the data from social media can help to better understand public emotions and concerns related to COVID-19 at the city level, which will potentially enable developing acceptable policies.

Acknowledgments

This work is supported in part by funds from the Medical Research Council (MRC), UK MR/R022461/1, and the Alan Turing Institute, United Kingdom.

Authors' Contributions

HA curated the data for this study. HA and TZ designed methodologies, analyzed and interpreted the data and drafted the manuscript. SA edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Percentages of tweets in the data set associated with each individual emoji.
[PDF File (Adobe PDF File), 178 KB - jmir_v24i10e40323_app1.pdf]

Multimedia Appendix 2
Examples of tweets expressing positive and negative reactions about COVID-19.
[PDF File (Adobe PDF File), 184 KB - jmir_v24i10e40323_app2.pdf]

References

1. Shuja J, Alanazi E, Alasmary W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Appl Intell 2021;51(3):1296-1325 [FREE Full text] [doi: 10.1007/s10489-020-01862-6] [Medline: 34764552]
2. Debata B, Patnaik P, Mishra A. COVID-19 pandemic! It's impact on people, economy, and environment. J Public Affairs 2020 Sep 02;20:e2372. [doi: 10.1002/pa.2372]
3. Hussain MW, Mirza T, Hassan MM. Impact of COVID-19 Pandemic on the Human Behavior. Int J Educ Manag Eng 2020 Dec 08;10(6):35-61. [doi: 10.5815/ijeme.2020.05.05]
4. Sparasci O, Bhui K, Biswas A, Chamberlain S, Dubicka B, Dudas R, et al. Impact of COVID-19 on mental health research: is this the breaking point? Br J Psychiatry 2022 Feb 17:1-3. [doi: 10.1192/bjp.2022.8] [Medline: 35172915]
5. Top 25 surprising Twitter statistics UK edition. cybercrew. 2022. URL: https://cybercrew.uk/blog/twitter-statistics-uk/ [accessed 2022-09-28]
6. Paul M, Dredze M. You are what you tweet: analyzing twitter for public health. 2011 Presented at: Fifth International AAAI Conference on Web and Social Media; July 17-21, 2011; Barcelona, Spain.
7. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant R. 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]
8. Ainley E, Witwicki C, Tallett A, Graham C. Using Twitter comments to understand people's experiences of UK health care during the COVID-19 pandemic: thematic and sentiment analysis. J Med Internet Res 2021 Oct 25;23(10):e31101 [FREE Full text] [doi: 10.2196/31101] [Medline: 34469327]
9. Zhang C, Xu S, Li Z, Hu S. Understanding concerns, sentiments, and disparities among population groups during the COVID-19 pandemic via Twitter data mining: large-scale cross-sectional study. J Med Internet Res 2021 Mar 05;23(3):e26482 [FREE Full text] [doi: 10.2196/26482] [Medline: 33617460]
10. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infect Control Hosp Epidemiol 2021 Feb;42(2):131-138 [FREE Full text] [doi: 10.1017/ice.2020.406] [Medline: 32758315]

11. Kleinberg B, van DVI, Mozes M. Measuring emotions in the COVID-19 real world Worry Dataset. 2020 Presented at: 1st Workshop on NLP for COVID-19 at ACL 2020; July 5-10, 2020; online.
12. Li X, Zhou M, Wu J. Analyzing COVID-19 on online social media: trends, sentiments and emotions. arXiv. 2020. URL: https://arxiv.org/abs/2005.14464 [accessed 2022-09-28]
13. Gupta RK, Vishwanath A, Yang Y. COVID-19 Twitter dataset with latent topics, sentiments and emotions attributes. arXiv. 2020. URL: https://arxiv.org/abs/2007.06954 [accessed 2022-09-28]
14. Glandt K, Khanal S, Li Y. Stance detection in COVID-19 tweets. 2021 Presented at: 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; August 1-6, 2021; Bangkok, Thailand. [doi: 10.18653/v1/2021.acl-long.127]
15. Kruspe A, Häberle M, Kuhn I. Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic. 2020 Presented at: 1st Workshop on NLP for COVID-19 at ACL 2020; July 5-10, 2020; online.
16. Hossain T, Logan IRL, Ugarte A. COVIDLies: Detecting COVID-19 misinformation on social media. 2020 Presented at: 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; December 2020; online. [doi: 10.18653/v1/2020.nlpcovid19-2.11]
17. 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]
18. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
19. 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]
20. 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]
21. Addawood A, Alsuwailem A, Alohali A. Tracking and understanding public reaction during COVID-19: Saudi Arabia as a use case. 2020 Presented at: 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; December 2020; online URL: https://doi.org/10.18653/v1/2020.nlpcovid19-2.24 [doi: 10.18653/v1/2020.nlpcovid19-2.24]
22. 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]
23. Gorodnichenko Y, Pham T, Talavera O. Social media, sentiment and public opinions: evidence from #Brexit and #USElection. Eur Econ Rev 2021 Jul;136:103772. [doi: 10.1016/j.euroecorev.2021.103772]
24. Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. JMIR Med Inform 2020 Jan 28;8(1):e16023 [FREE Full text] [doi: 10.2196/16023] [Medline: 32012057]
25. Zhang L, Hall M, Bastola D. Utilizing Twitter data for analysis of chemotherapy. Int J Med Inform 2018 Dec;120:92-100. [doi: 10.1016/j.ijmedinf.2018.10.002] [Medline: 30409350]
26. Andalibi N, Buss J. The human in emotion recognition on social media: attitudes, outcomes, risks. 2020 Presented at: CHI '20: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376680]
27. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med 2022 Apr 08;5(1):46. [doi: 10.1038/s41746-022-00589-7] [Medline: 35396451]
28. Ji S, Pan S, Li X, Cambria E, Long G, Huang Z. Suicidal ideation detection: a review of machine learning methods and applications. IEEE Trans Comput Soc Syst 2021 Feb;8(1):214-226. [doi: 10.1109/tcss.2020.3021467]
29. Weissenbacher D, Sarker A, Magge A. Overview of the Fourth Social Media Mining for Health (SMM4H) Shared Tasks at ACL 2019. 2019 Presented at: Fourth Social Media Mining for Health Applications (# SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 21-30. [doi: 10.18653/v1/w19-3203]
30. Aragón ME, Monroy APL, González-Gurrola LC. Detecting depression in social media using fine-grained emotions. 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2019; Minneapolis, MN p. 1481-1486 URL: https://doi.org/10.18653/v1/n19-1151 [doi: 10.18653/v1/n19-1151]
31. Chen X, Sykora MD, Jackson TW. What about mood swings? Identifying depression on twitter with temporal measures of emotions. 2018 Presented at: WWW '18: The Web Conference 2018; April 2018; Lyon, France p. 1653-1660. [doi: 10.1145/3184558.3191624]
32. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. J Biomed Inform 2016 Aug;62:148-158 [FREE Full text] [doi: 10.1016/j.jbi.2016.06.007] [Medline: 27363901]

33. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
34. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
35. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
36. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
37. 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]
38. Cambria E, Li Y, Xing FZ. SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. 2020 Presented at: CIKM '20: 29th ACM International Conference on Information & Knowledge Management; October 19, 2020; online. [doi: 10.1145/3340531.3412003]
39. Alhuzali H, Ananiadou S. SpanEmo: Casting multi-label emotion classification as span-prediction. 2021 Presented at: EACL 2021: 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume; April 19-23, 2021; online p. 1573-1584. [doi: 10.18653/v1/2021.eacl-main.135]
40. Bianchi F, Terragni S, Hovy D. Pre-training is a hot topic: contextualized document embeddings improve topic coherence. 2021 Presented at: 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; August 2021; online p. 759-766. [doi: 10.18653/v1/2021.acl-short.96]
41. geopy 2.2.0. URL: https://pypi.org/project/geopy/ [accessed 2022-09-28]
42. Welcome to GeoPy's documentation!. GeoPy. URL: https://geopy.readthedocs.io/en/stable/#nominatim [accessed 2022-09-28]
43. United Kingdom: Countries and Major Cities. City Population. URL: https://www.citypopulation.de/en/uk/cities/ [accessed 2022-09-28]
44. Baziotis C, Pelekis N, Doulkeridis C. DataStories at SemEval-2017 Task 4: deep LSTM with attention for message-level and topic-based sentiment analysis. 2017 Presented at: 11th International Workshop on Semantic Evaluation (SemEval-2017); August 2017; Vancouver, BC. [doi: 10.18653/v1/s17-2126]
45. Poria S, Cambria E, Gelbukh A, Bisio F, Hussain A. Sentiment data flow analysis by means of dynamic linguistic patterns. IEEE Comput Intell Mag 2015 Nov;10(4):26-36. [doi: 10.1109/mci.2015.2471215]
46. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. SemEval-2018 Task 1: Affect in Tweets. 2018 Presented at: 12th International Workshop on Semantic Evaluation; June 2018; New Orleans, LA. [doi: 10.18653/v1/s18-1001]
47. Blei DM, Ng AY, Jordan MI. Latent dirichllocation. J Mach Learn Res 2003;3:993-1022.
48. Benton A, Coppersmith G, Dredze M. Ethical research protocols for social media health research. 2017 Presented at: First ACL Workshop on Ethics in Natural Language Processing; April 2017; Valencia, Spain p. 94-102. [doi: 10.18653/v1/w17-1612]
49. Mohammad S. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. 2018 Presented at: 56th Annual Meeting of the Association for Computational Linguistics; July 2018; Melbourne, Australia p. 174-184. [doi: 10.18653/v1/p18-1017]
50. Xu P, Liu Z, Winata GI, Lin Z, Fung P. Emograph: Capturing emotion correlations using graph networks. arXiv. 2020. URL: http://128.84.4.34/abs/2008.09378 [accessed 2022-09-28]
51. Saif M, Kiritchenko S. Understanding emotions: a dataset of tweets to study interactions between affect categories. 2018 Presented at: Eleventh International Conference on Language Resources and Evaluation; May 2018; Miyazaki, Japan.
52. Barletta WA. The influence of SARS-CoV-2 variants on national case-fatality rates: correlation and validation study. JMIRx Med 2022;3(2):e32935 [FREE Full text] [doi: 10.2196/32935] [Medline: 35969709]
53. Shoeb AAM, de Melo G. EmoTag1200: Understanding the association between emojis and emotions. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 2020; online p. 8957-8967. [doi: 10.18653/v1/2020.emnlp-main.720]
54. Felbo B, Mislove A, Søgaard A. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. 2017 Presented at: Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen, Denmark p. 1615-1625. [doi: 10.18653/v1/d17-1169]
55. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social network analysis of COVID-19 sentiments: application of artificial intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]

Abbreviations

CTM: combined topic modeling

NHS: National Health Service

Edited by C Basch; submitted 15.06.22; peer-reviewed by J Nielsen, M Rabbani; comments to author 08.07.22; revised version received 06.08.22; accepted 10.08.22; published 05.10.22.

Please cite as:
Alhuzali H, Zhang T, Ananiadou S
Emotions and Topics Expressed on Twitter During the COVID-19 Pandemic in the United Kingdom: Comparative Geolocation and Text Mining Analysis
J Med Internet Res 2022;24(10):e40323
URL: https://www.jmir.org/2022/10/e40323
doi:10.2196/40323
PMID:36150046

Original Paper

Fine-tuned Sentiment Analysis of COVID-19 Vaccine–Related Social Media Data: Comparative Study

Chad A Melton[1,2], MSc; Brianna M White[2], MPH; Robert L Davis[2], MD, MPH; Robert A Bednarczyk[3], PhD; Arash Shaban-Nejad[2], MPH, PhD

[1]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee at Knoxville, Knoxville, TN, United States

[2]Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[3]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Arash Shaban-Nejad, MPH, PhD
Center for Biomedical Informatics
Department of Pediatrics, College of Medicine
University of Tennessee Health Science Center
50 N Dunlap Street, 492R
Memphis, TN, 38103
United States
Phone: 1 9012875836
Email: ashabann@uthsc.edu

Abstract

Background: The emergence of the novel coronavirus (COVID-19) and the necessary separation of populations have led to an unprecedented number of new social media users seeking information related to the pandemic. Currently, with an estimated 4.5 billion users worldwide, social media data offer an opportunity for near real-time analysis of large bodies of text related to disease outbreaks and vaccination. These analyses can be used by officials to develop appropriate public health messaging, digital interventions, educational materials, and policies.

Objective: Our study investigated and compared public sentiment related to COVID-19 vaccines expressed on 2 popular social media platforms—Reddit and Twitter—harvested from January 1, 2020, to March 1, 2022.

Methods: To accomplish this task, we created a fine-tuned DistilRoBERTa model to predict the sentiments of approximately 9.5 million tweets and 70 thousand Reddit comments. To fine-tune our model, our team manually labeled the sentiment of 3600 tweets and then augmented our data set through back-translation. Text sentiment for each social media platform was then classified with our fine-tuned model using Python programming language and the Hugging Face sentiment analysis pipeline.

Results: Our results determined that the average sentiment expressed on Twitter was more negative (5,215,830/9,518,270, 54.8%) than positive, and the sentiment expressed on Reddit was more positive (42,316/67,962, 62.3%) than negative. Although the average sentiment was found to vary between these social media platforms, both platforms displayed similar behavior related to the sentiment shared at key vaccine-related developments during the pandemic.

Conclusions: Considering this similar trend in shared sentiment demonstrated across social media platforms, Twitter and Reddit continue to be valuable data sources that public health officials can use to strengthen vaccine confidence and combat misinformation. As the spread of misinformation poses a range of psychological and psychosocial risks (anxiety and fear, etc), there is an urgency in understanding the public perspective and attitude toward shared falsities. Comprehensive educational delivery systems tailored to a population's expressed sentiments that facilitate digital literacy, health information–seeking behavior, and precision health promotion could aid in clarifying such misinformation.

(J Med Internet Res 2022;24(10):e40408) doi:10.2196/40408

KEYWORDS

sentiment analysis; DistilRoBERTa; natural language processing; social media; Twitter; Reddit; COVID-19; vaccination; vaccine; content analysis; public health; surveillance; misinformation; infodemiology; information quality

Introduction

Background

The novel coronavirus (COVID-19) has impacted and disrupted many aspects of everyday life worldwide. Following the implementation of rigid pandemic mitigation strategies in early 2020, social media use substantially increased with internet users turning to social media platforms to communicate and gather information regarding the dynamic and uncertain situation [1-4]. As the pandemic progressed and researchers worked to develop vaccines, many social media users turned their focus to gathering information regarding various topics related to COVID-19 vaccines, such as side effects, availability, and efficacy. As of May 19, 2022, approximately 6.27 million people across the world have died due to complications from COVID-19. Moreover, many experience long COVID syndrome, in which viral symptoms persist past the expected clinical recovery time [5]. Although COVID-19 vaccines are safe and effective at preventing life-threatening infections, hospitalizations, and deaths, vaccine hesitancy related to COVID-19 vaccines has led to further comorbidities and many preventable deaths [6-8].

With an estimated 4.5 billion users worldwide, social media offers an opportunity for near real-time analysis of large bodies of text data (500 million tweets/day) that could be useful to public health officials [3,9]. Using machine/deep learning, recent advancements in natural language processing methods (eg, Bidirectional Encoder Representations from Transformers [BERT], RoBERTa, GPT2, and XLNet) have substantially improved previous text classification models (greater than 90% accuracy) [4,10-14]. Moreover, pretrained models such as BERT or RoBERTa are available and free to researchers from platforms such as Hugging Face. These platforms are extremely helpful to the greater scientific community, considering that many of these models take several days on dozens of tensor processing units to learn [15,16]. Importantly, these models can be fine-tuned based on a particular use case (eg, text classification, text generation, and sentiment analysis). The enhanced functionality provides a researcher with techniques to investigate a wide variety of phenomena across many scientific domains [17-19]. Sentiment analysis (ie, classifying text as positive or negative) in particular is a powerful tool that can be used to correlate events to the public mood, surveil public health discussion, and even detect disease outbreaks [18]. Most importantly, these methods can be used by public health officials to develop precise messaging strategies and intervention campaigns to address the information crises and improve vaccination rates.

Our study sought to examine and explore sentiment regarding COVID-19 vaccines expressed on 2 popular social media platforms—Reddit and Twitter. We calculated positive and negative sentiment by creating a custom fine-tuned DistilRoBERTa model with data labeled by members of our team and then augmented by back-translation. We then offered a comparison of sentiment regarding COVID-19 vaccines across Reddit and Twitter. We hypothesized that we would observe somewhat similar trends in polarity between the 2 social media platforms with minor differences, because DistilRoBERTa has typically displayed accuracies greater than 90% [16]. However, we expected that our labeled data set would provide more nuanced insight into public sentiments in these 2 communities than previous sentiment analysis methods. Additionally, based on our previous work, we hypothesized that sentiment would remain more positive than negative [4]. Finally, we argued that identifying and following social media shared sentiment allows for the eventual development of comprehensive response strategies, which are better aimed at combatting misinformation and disinformation; improvement of vaccine delivery; and containment of disease transmission.

COVID-19–Related Social Media Analysis

Social media content analysis is not a brand new concept and has been used for data mining and sentiment analysis before COVID-19. However, the nature of the pandemic response and the necessary separation of populations for safety have led to an unprecedented number of new users [9]. This influx caused a surge in social networking posts, leaving researchers with mountains of content to sort through. One positive aspect of social media data mining is that the content is publicly available and easily obtainable, allowing for rapid collection. The rapid collection of data, especially those related to COVID-19, permits researchers to follow the pandemic's progression alongside sentiment on the web. For example, the ability to rapidly collect tweets from a specified time period allows for the parallel analysis of general public opinion during major events, such as the release of the Pfizer vaccine in late 2020 or the death of a celebrity post–COVID-19 infection [20]. This targeted approach provides tools for niche discovery and exploration of the sentiment behind health decision-making.

Researchers have used the recent increase in opinion sharing to measure overall sentiment and vaccine hesitancy or acceptance [4,20-24]. As social media usage has continued to grow throughout the pandemic era, more than 3.6 billion people are known to regularly log on to at least one networking platform. Twitter is considered one of the largest and most used social media platforms, with more than 400 million account owners [9]. The platform allows users to post short messages or tweets for "followers" to see and respond to, based on the underlying sentiment they evoke. Tweets are limited to brief messages, with a 280-character limit, but may contain attached images, videos, or highlighted popular keywords known as "hashtags." Additionally, tweets can include hyperlinks to news articles or scientific literature. If another user agrees with a posted tweet, they can "retweet" or share the message to their profiles in a show of rapport. Rather than joining topic-based communities, users typically follow other users.

The Reddit platform is similar in size, with approximately 430 million current users [9]. However, it is different in message format and delivery, in that users are allowed to create groupings based on a topic, called "subreddits." Subreddits often contain open dialogue alongside images, videos, and hyperlinks to news articles or literature. Similar to "retweeting," subreddit subscribers have the unique ability to "upvote" or "downvote" a post based on the user's opinion of its contents. Users are also able to join the discussion by leaving comments, which can also

be upvoted or downvoted. If a subreddit becomes increasingly popular and receives a good share of upvotes, the post will appear first within a topic category. The more traffic a subreddit receives, even if it is sharing misinformation or disinformation, the higher the Reddit platform will promote it. Notably, subreddits generally have rules that community members must adhere to or risk the potential for the removal of a post or banning.

BERT Algorithm

Substantial advances in natural language processing have occurred since the development of BERT and the work built from its architecture. BERT is a powerful and versatile artificial intelligence–based natural language processing algorithm developed at Google AI Language that excels at text classification (ie, ontologies, categories, and sentiment, etc) of unstructured/semistructured text data that are characteristic of social media data [10]. The BERT algorithm was trained on the entirety of Wikipedia and the Brown Corpus over 4 days using 16 cloud-based tensor processing units. BERT is a transformer-based language model that uses multiple encoders to create word embeddings. These embeddings are then used in concert with masked language modeling and next sentence prediction to learn by predicting random masked words in a sentence and learning to predict sentences, respectively. These 2 steps teach BERT to understand context, a skill that older recurrent neural networks typically struggled with. A convenient aspect of BERT is that it has the capability to fine-tune the model with relevant data by replacing the output layer with weights from custom data. Researchers have been inspired by the original BERT architecture to create many variations (eg, RoBERTa, DistilRoBERTa, DistilBERT, and BART, etc) that have surpassed the benchmarks of previous models. Moreover, these models can be fine-tuned for specific domain-based tasks (ClinicalBERT and BioBERT) in multiple languages [11,12,25]. Furthermore, several studies have used other fine-tuned BERT models to investigate COVID-19–related content expressed on social media related to misinformation detection, sentiment classification, and continent analysis [13,26-29].

Methods

Study Overview

Our study compared COVID-19 vaccine–related postings from 2 popular social media platforms—Reddit and Twitter—from January 1, 2020, to March 1, 2022. These 2 platforms were chosen due to their worldwide usage, vibrant discussions, and high user count. The time frame included the earliest parts of the pandemic to trace the evolution of sentiments over time. Most importantly, these platforms were chosen because only a small number of comparative studies have focused on the typical user, especially studies related to COVID-19 vaccine sentiment or other vaccines. Our study used a binary (ie, positive or negative polarity) sentiment classification method for training our model and for sentiment analysis. A binary system was chosen for a few reasons. (1) Binary systems are more computationally efficient when processing large bodies of data. (2) Binary classifiers are typically more accurate than multiclass systems. (3) In the past, sentiment classifiers that incorporate a neutral class often rely on a low probability or confidence score. Since our model reported a confidence value, this information could be extrapolated.

Data Overview

Substantial effort was taken to identify and remove Twitter posts that were found to be directly from news agencies or bots. These posts were identified by their source having an overwhelmingly high post count during the 26-month period relative to the average number of posts of a “normal” user, as well as by visually inspecting tweets of users that appeared at an abnormal frequency. Both Twitter and Reddit data sets were limited to only include users who posted fewer than or equal to 200 times throughout our time frame. These steps were important due to the repetitive nature of many bot tweets, which had the potential to skew sentiment calculations and misalign the goal to compare the normal user base of both platforms. Although the methodologies in harvesting Reddit and Twitter data differ slightly, both data sets underwent similar cleaning steps. Both data sets were queried for the same relevant terms typically present in web-based discussions about COVID-19 vaccines. This step was important due to the tendency for some extended comment threads to meander off-topic. This occurrence was especially true with threads from some Reddit communities. The daily posting frequencies of the 2 platforms were relatively similar in the early months of the pandemic. The frequency increased dramatically for both platforms in late September to October 2020 as news of vaccine circulation became more widespread. Although each platform displayed 4 spikes in posting frequency at similar time periods (October 2020, March to April 2021, August to September 2021, and December 2021 to January 2022), they obtained a maximum in different time periods. Reddit reached its maximum posting from March to April 2021, whereas Twitter reached its maximum from September to October 2021.

Twitter

Approximately 13 million tweets were harvested using the *snscrape* and *Tweepy* API Python libraries based on the search term “COVID Vaccine.” After removing tweets by suspected bots, news media, or highly repetitive high-frequency users and duplicate tweets, our final Twitter data set consisted of 9,518,270 tweets authored by 3,006,075 Twitter users. The tweets contained approximately 16.32 million total likes, with a maximum of 430,758 likes and an average of 14.9 likes per tweet. Tweets cannot be downvoted, but approximately 4,794,865 tweets were attributed with 0 likes. Statistics on tweet sharing or retweets were not collected because this metric was not available for both platforms.

Reddit

We harvested 579,241 user-created posts from 67 subreddits with the Python Reddit API *Wrapper*. These subreddits were collected to gain a broad understanding of sentiments related to the COVID-19 vaccines as well as to avoid potential biases in data collection. These subreddits contained a total of 5,590,913 subscribers as of March 1, 2022. Our query removed a large portion of unrelated terms. After visually inspecting and confirming the results of the querying process, our final Reddit

data set consisted of 67,962 comments composed by at least 9843 authors. These posts contained approximately 2.1 million total upvotes, with an average of 31 upvotes and a maximum of 18,253 upvotes per comment.

Data Labeling and Augmentation

Since time is of the essence in a global pandemic, combined with the fact that labeling data is time-consuming and costly, we created a custom training data set by labeling sentiment (positive or negative) for approximately 3600 tweets related to COVID-19 vaccines. We chose to label tweets exclusively for this study, because the 280-character limit of a tweet (ie, compared to a Reddit post limit of a maximum of 10,000 characters) would allow our small team to create a time-relevant training data set more quickly. We then augmented our data set through the process of back-translation with several language models on the Hugging Face model repository. Back-translation was chosen after testing a few other methods of text augmentation. Some techniques (eg, word masking) resulted in far more duplicated texts that would eventually need to be removed. Back-translation relies on subtle differences between language structure, word meaning, and syntax. In effect, the outputted text will vary slightly from the inputted text without losing semantic and contextual meaning [14]. In our case, the back-translation method translated our English-language text into another language (eg, French, Chinese, Greek, and Hebrew) and then back into English. After removing duplicates, our final augmented data set consisted of 48,691 tweets.

RoBERTa and DistilRoBERTa

For our study, we chose to explore the capabilities of DistilRoBERTa. RoBERTa is a more robust model than BERT, and DistilRoBERTa is an optimized version of RoBERTa [15,16]. Developed at Facebook, RoBERTa was trained on 160 GB of text compared to the 16 GB of BERT. RoBERTa dropped the next sentence prediction feature of BERT and added dynamic token masking during training. These enhancements are estimated to have improved the original BERT's performance significantly (2% to 20%) [16]. Compared to RoBERTa, DistilRoBERTa was trained on approximately 40 GB of text data (OpenWebTextCorpus) and operates about twice as fast.

The University of Tennessee Health Science Center Vaccine Sentiment Labeling and DistilRoBERTa Fine-tuning

We fine-tuned the DistilRoBERTa base via the Hugging Face *Trainer* class, which provides the user with an API for training with *PyTorch*. Our data were then randomized and segregated into 40,000 training tweets, 4000 validation tweets, and 4691 tweets for testing. Training hyperparameters included a 2×10^{-5} learning rate, 32 training and evaluation batch size, a seed number of 42, and a linear scheduler with 500 warm-up steps. We used the *Adam* optimizer with betas of 0.9 and 0.999 and an epsilon of 1×10^{-8}. Lastly, our model was trained for 2 epochs. These hyperparameters achieved a training loss of 0.1284, a validation loss of 0.1167, a precision of 0.9561, an F_1-score of 0.9592, and an accuracy of 0.9592 (see Table 1).

Table 1. DistilRoBERTa fine-tuning training metrics. The model obtained optimal fine-tuning after 2 training epochs.

Step	Epoch	Training loss	Validation loss	Precision	Accuracy	F_1-score
500	0.4	0.5903	0.4695	0.7342	0.7728	0.7890
1000	0.8	0.3986	0.3469	0.8144	0.8596	0.8684
1500	1.2	0.2366	0.1939	0.9313	0.9260	0.9253
2000	1.6	0.1476	0.1560	0.9207	0.9452	0.9465
2500	2.0	0.1284	0.1167	0.9561	0.9592	0.9592

Analytical Methods

Following the fine-tuning of our model, we processed the Twitter and Reddit data through the Hugging Face *pipeline* for sentiment analysis. The model returned a label of either positive or negative for each tweet or Reddit comment. Along with the determined polarity, the model also returned a probabilistic confidence score ranging from 0 to 1. For clarity, tweets or comments classified as negative were multiplied by –1 to reflect the negative sentiment.

Ethical Considerations

No ethical approval was needed from our institution due to the public availability and nonidentifiable nature of the data used.

Results

DistilRoBERTa Fine-tuned to COVID-19 Vaccine

Twitter

The DistilRoBERTa fine-tuned polarity analysis determined that the 9,518,270 tweets were more negative (n=5,215,830, 54.8%) than positive (n=4,302,440, 45.2%) throughout our time frame (see Figure 1).

The maximum positive rating occurred in March 2021 (375,789/675,274 55.6%). However, the minimum positive rating occurred in January 2022 (191,159/526,582, 36.3%), displaying a steady decrease in polarity from the maximum. For the confidence score, the tweets classified as positive had a maximum score of 0.999, a minimum of approximately 0 (3.58×10^{-7}), and a mean of 0.868 (see Figure 2). The tweets classified as negative had a minimum score of –0.999, a

maximum value of approximately zero (-1.78×10^{-6}), and a mean of –0.882 (see Figures 1 and 2).

Figure 1. Tweet polarity from the DistilRoBERTa model fine-tuned to COVID-19 vaccine. Polarity and the corresponding confidence probability are represented on the y-axis, and time is represented on the x-axis. Tweets are represented as light blue circles. Circle size indicates the number of likes per tweet—larger circles indicate more likes and smaller circles indicate fewer likes.

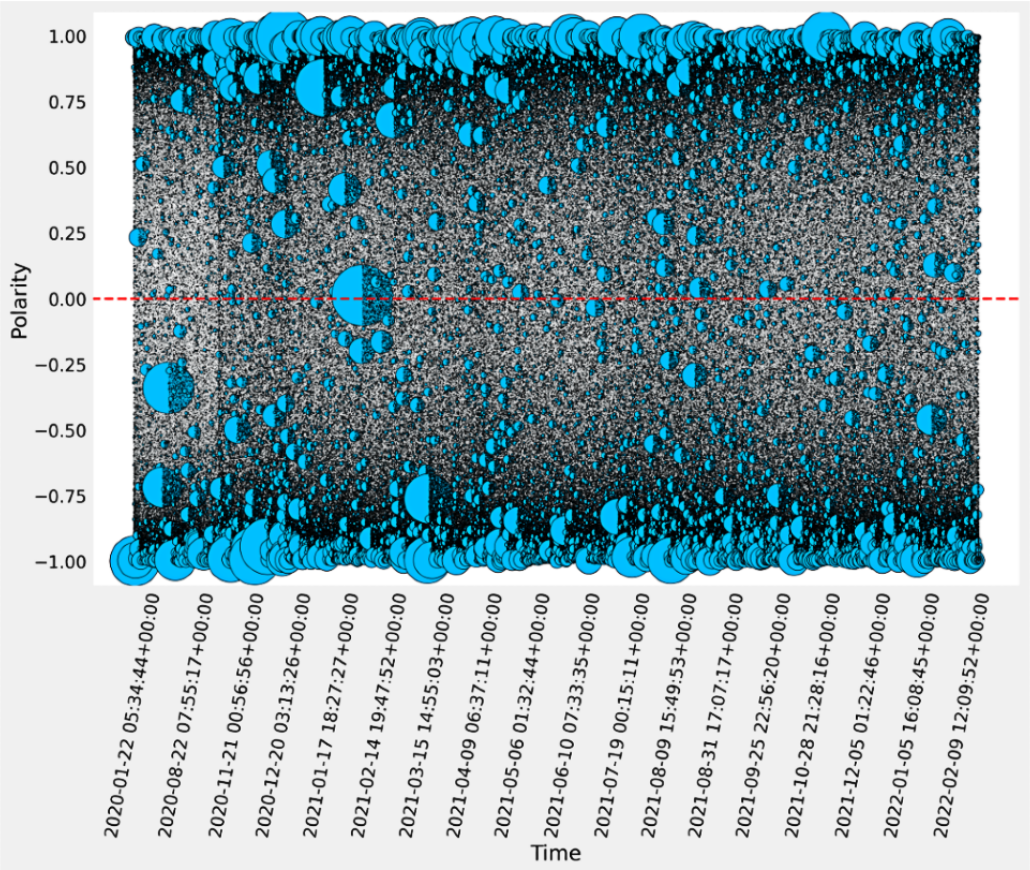
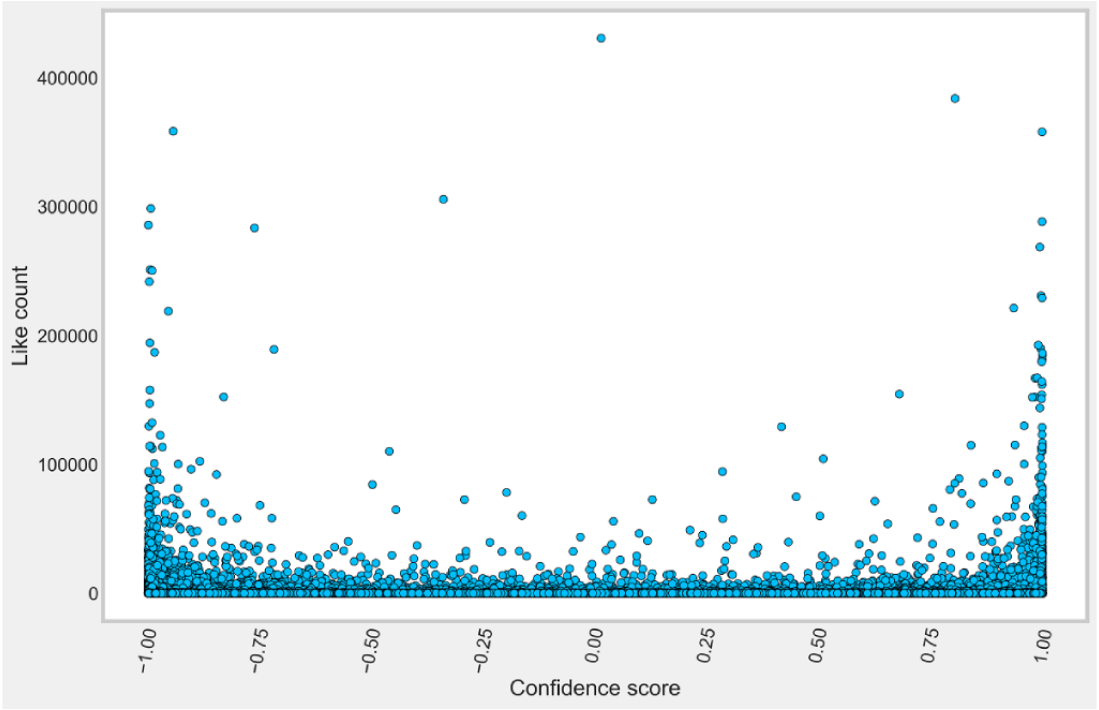


Figure 2. Confidence score versus like count for Twitter. The x-axis represents the confidence score and the y-axis represents the number of likes a tweet received. Data points below 0.00 on the x-axis represent a negative classification, and data points above 0.00 represent a positive classification. Data points are represented as light blue circles.



Reddit

The Reddit sentiment polarity analysis for the fine-tuned DistilRoBERTa model found that of the 67,962 posts, 37.7% (n=25,646) were classified as negative and 62.3% (n=42,316) were classified as positive. The highest polarity reported in our experiment and the maximum positive rating occurred in April 2021 (6611/9044, 73.1 %), and the minimum positive rating occurred in February 2020 (170/351, 48.4%). For the confidence scores, the comments classified as positive had a maximum score of 0.999, a minimum of approximately 0 (1.55×10^{-4}), and a mean of 0.870 (see Figure 3). The comments classified as negative had a minimum of –0.999, a maximum of approximately 0 (-4.74×10^{-5}), and a mean of –0.808 (see Figures 3 and 4).

Figure 3. Reddit comment polarity from the DistilRoBERTa model fine-tuned to COVID-19 vaccine. Polarity and corresponding confidence probability are represented on the y-axis, and time is represented on the x-axis. Data points are represented as orange-red circles. Circle size indicates the number of upvotes per comment—more upvotes are represented by larger circles and fewer upvotes are represented by smaller circles.

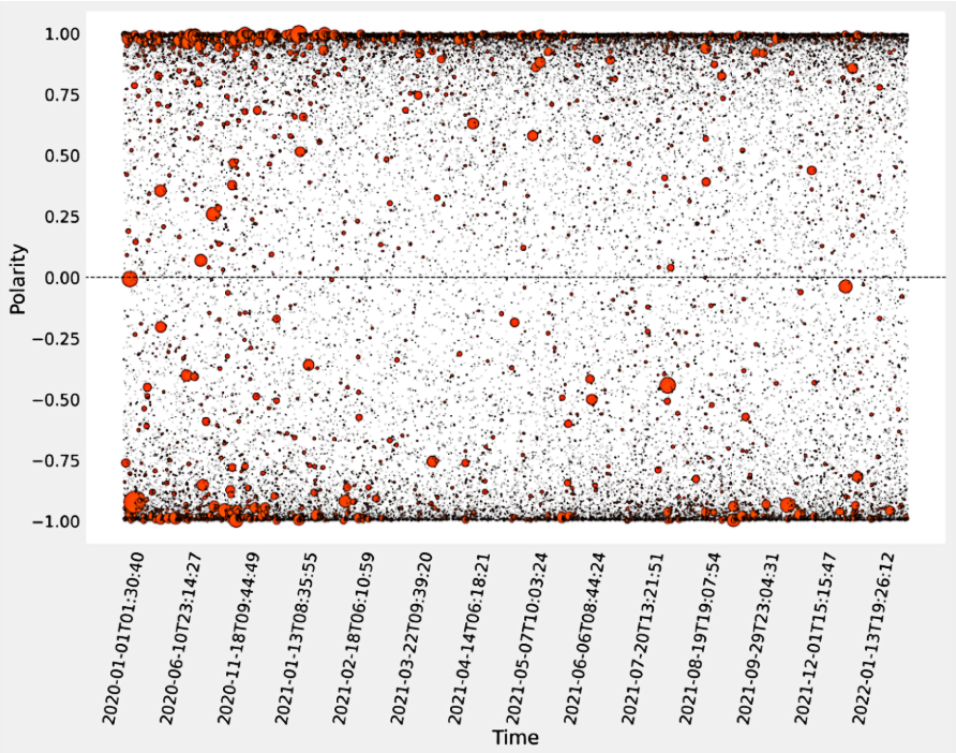
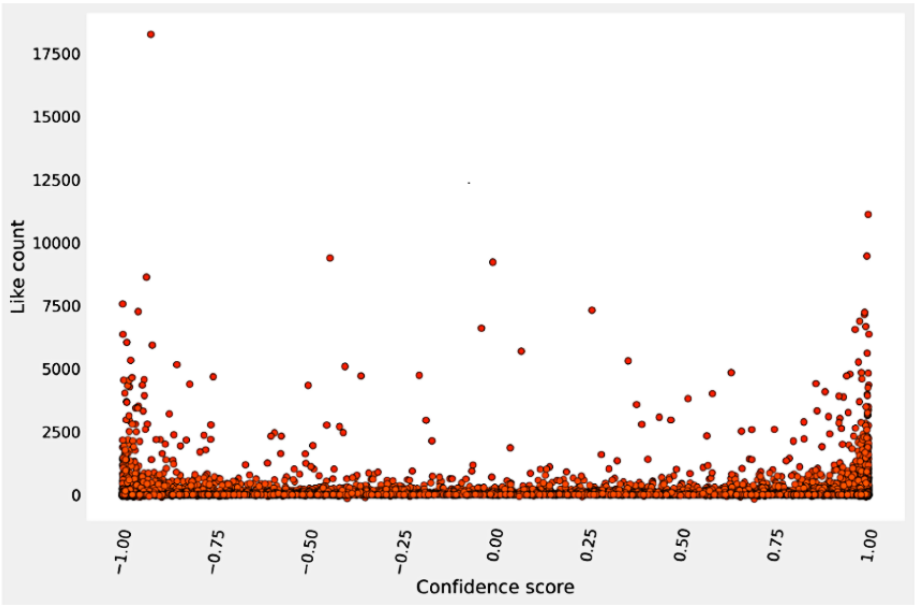


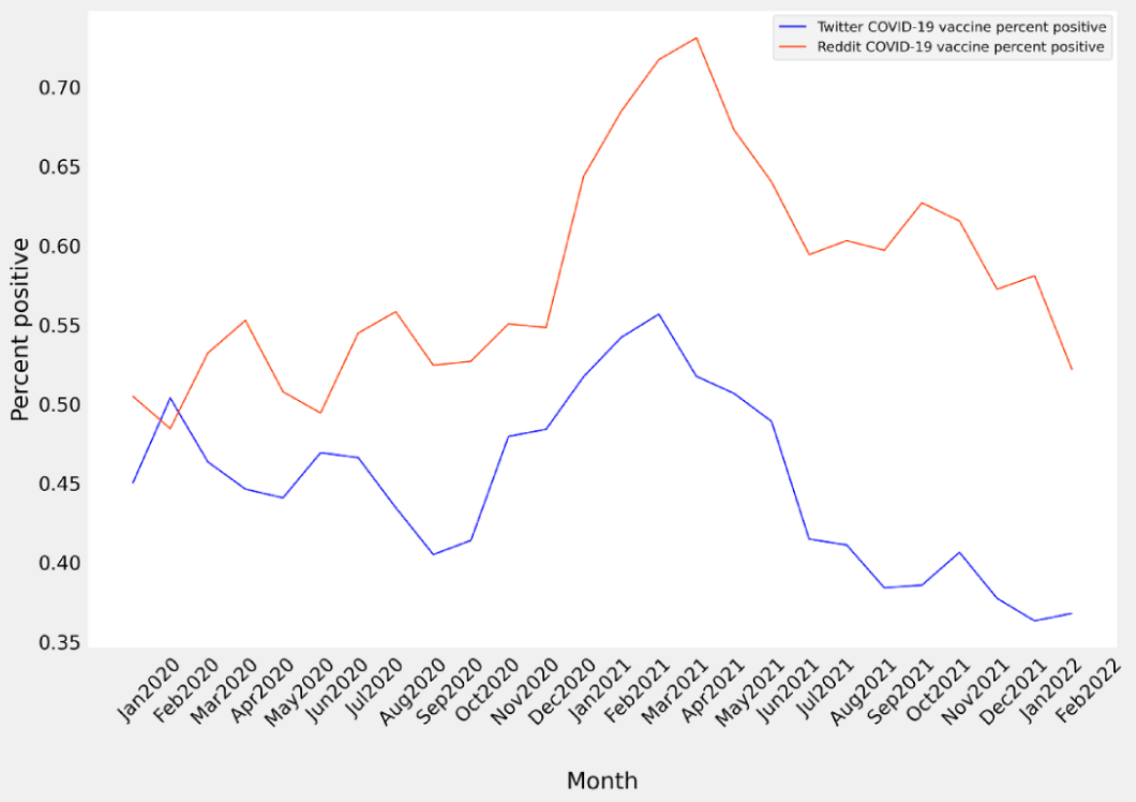
Figure 4. Confidence score versus like count for Reddit. The x-axis represents the confidence score and the y-axis represents the number of upvotes a comment received. Data points below 0.00 on the x-axis represent a negative classification, and data points above 0.00 represent a positive classification. Data points are represented as orange-red circles.



COVID-19 Vaccine Sentiment Expressed on Reddit and Twitter

Overall, the average sentiment for the 2 social media platforms was somewhat different (62.3% positive on Reddit vs 45.2% positive on Twitter). An interesting story begins to appear when looking closely at the month-to-month results in relation to each other. Although sentiment on both platforms oscillated in the early months of the pandemic, Reddit sentiment was higher (ranging from 48% to 55% positive) from January to August 2020. Twitter sentiment began similar to Reddit sentiment but gradually declined until becoming substantially more negative from September to October 2020, and then increasing to a maximum of 55% in March 2021. Reddit sentiment began a steep increase in polarity in December 2020 and continued to increase until reaching the maximum positive sentiment (approximately 73%) in April 2021. After sentiment on each platform achieved their maximum positive polarity, both began an oscillating and gradual decline in sentiment to near early pandemic levels. However, Twitter sentiment continued to fall until achieving a minimum of 36% (see Figure 5).

Figure 5. Monthly sentiment for Twitter and Reddit COVID-19 vaccine–related posts. The x-axis represents time and the y-axis represents the percentage of posts classified as positive. The blue line represents Twitter sentiment and the orange-red line represents Reddit sentiment. Note that since posting frequency was very low, sentiment for January 2020 is an average of all other months of corresponding data.



Discussion

Interpretation of Results

Ranging from January 1, 2020, to March 1, 2022, our results show that the average sentiment for the Reddit data set was more positive than the average sentiment expressed on Twitter. Interestingly, both platforms expressed similar sentiment changes during key moments of the pandemic (eg, vaccine efficacy announcements, vaccine distribution to all ages, new variants, and waning efficacy). This behavior is especially observable as vaccines became widely available to the public and the polarity diminished. Considering this similar behavior, we feel that both Twitter and Reddit continue to be valuable data sources that public health officials can use to develop vaccine education campaigns and digital interventions. Although Twitter is superior in the ability to access large numbers of tweets through an API, substantial steps need to be taken while cleaning Twitter data to remove bots, news media posts, commercial users, duplicates, and users who have extremely high posting frequencies. On the other hand, Reddit data are more plentiful in longer texts that could be more useful for topic modeling.

What drove sentiment changes related to COVID-19 vaccines on these 2 platforms? One possibility could be related to the character limit of tweets versus Reddit posts (ie, 280 vs 10,000 characters, respectively). The shortened character limit of tweets most likely contributes to the quick spread of information and can be reactionary in nature, driving negative sentiment. However, Reddit users typically take advantage of the longer character limit and share, at times, highly personal stories and experiences related to their health care. For this reason, Reddit could remain a highly valuable source when considering the development of public health messaging and education campaigns.

Correlating changes in sentiment with developments during the pandemic presents some interesting challenges and ideas alike. The most obvious steep increase in sentiment seems to be correlated with positive news regarding vaccine development

and trials and news of high efficacy, distribution, and availability to those who patiently waited for the vaccine. It is challenging to correlate minimum sentiment scores because their decline was not uniform. It is highly likely that the gradual decline was related to a combination of unfortunate events related to the pandemic (eg, misinformation, pandemic fatigue, and falling vaccine efficacy). It is conceivable that challenges in vaccine rollout and distribution could negatively affect sentiment. However, previous topic modeling and semantic network analysis on portions of this data set did not find a meaningful occurrence of terms related to vaccine distribution. Therefore, more psychological, sociological, and cultural studies are desperately needed to understand what drives certain populations, news media, politicians, and entertainers to so readily accept and propagate misinformation and conspiracy theories rather than directly observable facts. Such studies would not only benefit future public health responses but also many other areas of life where misinformation and disinformation have taken hold. The success of digital interventions and education campaigns would likely be limited without a more thorough understanding of how to reach these populations.

Public Health Implications

The application of our findings could have momentous impacts on the public health sector in the fight against infectious diseases such as COVID-19. Further development of low–human effort surveillance systems optimized for the rapid collection of data would allow for the real-time analysis of public emotion in correlation with disease progression. Moreover, fine-tuning models to assess geographical and demographical differences in sentiment could provide insight into the attitudes of populations at the greatest risk of debilitating outcomes. In addition to geographically and demographically specific data mining, targeting public discourse during times of peak infection, vaccine releases, or the death of a celebrity, athlete, or political figure due to the disease could greatly bolster public health response [30,31]. The expansion of such disease projection and prediction models using sentiment mining techniques could also influence evidence-informed policy. Discerning the dynamic levels of population sentiment allows public health officials to design catered policy communication strategies. By providing the necessary tools to better understand public emotion related to disease prevention, control, and containment, policy makers would be better equipped to evaluate program successes and highlight any need for repositioning.

Furthermore, the analysis of sentiment shared via social media could prove to be a vital instrument in combatting rampant misinformation and disinformation shared on the web. As the spread of misinformation poses a range of psychological and psychosocial risks (anxiety and fear, etc), there is an urgency in understanding the public perspective and attitude toward shared falsities. Education delivery systems tailored to population-expressed sentiment could aid in clarifying such misinformation. Moreover, there is room for the expansion of artificially intelligent messaging systems, tasked with generating responses to waves of misinformation and disinformation shared via social media platforms. Overall, the proposed framework for the real-time analysis of sentiment could be useful in guiding governmental support of public health recovery efforts.

Limitations

As with most studies, ours has some limitations. Challenges occur when conducting sentiment analysis in social media texts due to some long-standing problems. Although BERT and newer models greatly mitigate many of these challenges, some models typically struggle with detecting sarcasm, humor, emotion, and complex inferences in texts unless specifically having been trained to do so. For example, many pro-vaccine social media users express extremely negative views and sentiments regarding the anti-vaccine community. How would BERT classify such an occurrence? Although their expressed sentiment is positive toward the vaccine, many natural language processing algorithms and data labelers would potentially struggle with this type of classification. Even though we took great care with this study to remove tweets by bots or tweets from highly repetitive users from Twitter and choose unbiased subreddits, it is possible that some could have still slipped through the data cleaning process. Moreover, augmented data can potentially cause problems with overfitting when fine-tuning models due to relatively similar semantic content. We limited our training epochs and closely monitored the relationship between training loss and validation loss to mitigate this potential problem. Future work could involve efforts to create a larger labeled data set that would include not only COVID-19 vaccine sentiments but those of other vaccines as well.

Conclusions

We conducted a sentiment analysis of approximately 70,000 Reddit comments and 9.5 million tweets with a fine-tuned DistilRoBERTa model. Our analysis found that both Reddit and Twitter users expressed similar changes in sentiment throughout the pandemic, even though Twitter was substantially more negative than Reddit. Although subtle differences in sentiment were observed monthly, both platforms demonstrated a substantial increase in positive sentiments as the COVID-19 vaccine became readily available to the general public. The results we present here are a portion of an ongoing study to investigate vaccine-related content on social media with a focus on identifying and combating misinformation in efforts to decrease vaccine hesitancy. Correlating strong sentiment with high infectivity rates could provide officials with forecasting for the public acceptance of migration strategies such as vaccine delivery and uptake. These integrated disease surveillance tools should not only be leveraged in the fight against COVID-19 but stand to play essential roles in the evolution of future health policy, decision-making, program implementation, and precision health promotion [32]. In the near future, our team plans to expand the methods demonstrated in this study into sentiment related to other types of vaccines (eg, human papillomavirus vaccines). We expect these results along with others to be used to develop tools to assist public health officials in monitoring public discourse regarding disease outbreaks, gaining a better understanding of vaccine hesitancy, and developing personalized digital interventions [33,34] and education campaigns.

Acknowledgments

We would like to thank our team of data labelers from the University of Tennessee Health Science Center. This study is partially supported by a grant (1R37CA234119-01A1) from the National Cancer Institute (NCI).

Data Availability

The data that support our findings are available upon reasonable request to the authors. Data are not available for commercial use.

Authors' Contributions

CAM conceptualized and supervised the study and drafted, reviewed, and edited the manuscript. BMW conceptualized the study and drafted, reviewed, and edited the manuscript. RLD reviewed and edited the manuscript. RAB reviewed and edited the manuscript. ASN drafted, reviewed, and edited the manuscript; supervised the study; and acquired funding.

Conflicts of Interest

None declared.

References

1. Sharma C, Whittle S, Haghighi PD, Burstein F, Keen H. Sentiment analysis of social media posts on pharmacotherapy: a scoping review. Pharmacol Res Perspect 2020 Oct 19;8(5):e00640 [FREE Full text] [doi: 10.1002/prp2.640] [Medline: 32813329]
2. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
3. 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 2022-03-20]
4. 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]
5. Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: an overview. Diabetes Metab Syndr 2021 May;15(3):869-875 [FREE Full text] [doi: 10.1016/j.dsx.2021.04.007] [Medline: 33892403]
6. Rosenberg ES, Holtgrave DR, Dorabawila V, Conroy M, Greene D, Lutterloh E, et al. New COVID-19 cases and hospitalizations among adults, by vaccination status - New York, May 3-July 25, 2021. MMWR Morb Mortal Wkly Rep 2021 Sep 17;70(37):1306-1311 [FREE Full text] [doi: 10.15585/mmwr.mm7037a7] [Medline: 34529645]
7. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2022-03-20]
8. Benham JL, Atabati O, Oxoby RJ, Mourali M, Shaffer B, Sheikh H, et al. COVID-19 vaccine-related attitudes and beliefs in Canada: national cross-sectional survey and cluster analysis. JMIR Public Health Surveill 2021 Dec 23;7(12):e30424 [FREE Full text] [doi: 10.2196/30424] [Medline: 34779784]
9. Clement J. Most popular social networks worldwide as of January 2022, ranked by number of monthly active users. Statista. 2022 Jan. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2022-03-20]
10. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. 2019 Jun Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
11. Alsentzer E, Murphy J, Boag WH, Weng W, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings. 2019 Jun Presented at: 2nd Clinical Natural Language Processing Workshop; June 7, 2019; Minneapolis, MN p. 72-78. [doi: 10.18653/v1/w19-1909]
12. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
13. Kolluri N, Liu Y, Murthy D. COVID-19 misinformation detection: machine-learned solutions to the infodemic. JMIR Infodemiology 2022 Aug 25;2(2):e38756. [doi: 10.2196/38756]
14. Beddiar DR, Jahan MS, Oussalah M. Data expansion using back translation and paraphrasing for hate speech detection. Online Soc Netw Media 2021 Jul;24:100153. [doi: 10.1016/j.osnem.2021.100153]
15. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. Roberta: a robustly optimized bert pretraining approach. ArXiv 2019 Jul 26:1-13. [doi: 10.48550/arXiv.1907.11692]
16. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv 2019 Oct 02:1-5. [doi: 10.48550/arXiv.1910.01108]

17. Blane JT, Bellutta D, Carley KM. Social-cyber maneuvers during the COVID-19 vaccine initial rollout: content analysis of tweets. J Med Internet Res 2022 Mar 07;24(3):e34040 [FREE Full text] [doi: 10.2196/34040] [Medline: 35044302]
18. Singh R, Singh R, Bhatia A. Sentiment analysis using machine learning technique to predict outbreaks and epidemics. Int J Adv Sci Res 2018 Mar;3(2):19-24 [FREE Full text]
19. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: a systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
20. Jennings W, Stoker G, Bunting H, Valgarðsson VO, Gaskell J, Devine D, et al. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. Vaccines (Basel) 2021 Jun 03;9(6):593 [FREE Full text] [doi: 10.3390/vaccines9060593] [Medline: 34204971]
21. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
22. Shakir A, Arora J. Accuracy in binary, ternary and multi-class classification sentimental analysis-a survey. Int J Advanced Res Comput Sci 2018 Feb 20;9(2):524-526. [doi: 10.26483/ijarcs.v9i2.5866]
23. Hou Z, Tong Y, Du F, Lu L, Zhao S, Yu K, et al. Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: a global social listening study. J Med Internet Res 2021 Jun 11;23(6):e27632 [FREE Full text] [doi: 10.2196/27632] [Medline: 34061757]
24. Kummervold PE, Martin S, Dada S, Kilich E, Denny C, Paterson P, et al. Categorizing vaccine confidence with a transformer-based machine learning model: analysis of nuances of vaccine sentiment in Twitter discourse. JMIR Med Inform 2021 Oct 08;9(10):e29584 [FREE Full text] [doi: 10.2196/29584] [Medline: 34623312]
25. Yum Y, Lee JM, Jang MJ, Kim Y, Kim J, Kim S, et al. A word pair dataset for semantic similarity and relatedness in Korean medical vocabulary: reference development and validation. JMIR Med Inform 2021 Jun 24;9(6):e29667 [FREE Full text] [doi: 10.2196/29667] [Medline: 34185005]
26. Ayoub J, Yang XJ, Zhou F. Combat COVID-19 infodemic using explainable natural language processing models. Inf Process Manag 2021 Jul;58(4):102569 [FREE Full text] [doi: 10.1016/j.ipm.2021.102569] [Medline: 33776192]
27. Koren A, Alam MAU, Koneru S, DeVito A, Abdallah L, Liu B. Nursing perspectives on the impacts of COVID-19: social media content analysis. JMIR Form Res 2021 Dec 10;5(12):e31358 [FREE Full text] [doi: 10.2196/31358] [Medline: 34623957]
28. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
29. Cresswell K, Tahir A, Sheikh Z, Hussain Z, Domínguez Hernández A, Harrison E, et al. Understanding public perceptions of COVID-19 contact tracing apps: artificial intelligence-enabled social media analysis. J Med Internet Res 2021 May 17;23(5):e26618 [FREE Full text] [doi: 10.2196/26618] [Medline: 33939622]
30. Salali GD, Uysal MS. Effective incentives for increasing COVID-19 vaccine uptake. Psychol Med 2021 Sep 20:1-3 [FREE Full text] [doi: 10.1017/S0033291721004013] [Medline: 34538287]
31. Romaniuc R, Guido A, Mai N, Spiegelman E, Sutan A. Increasing vaccine acceptance and uptake: a review of the evidence. SSRN Preprint posted online on May 11, 2021. [doi: 10.2139/ssrn.3839654]
32. Shaban-Nejad A, Michalowski M, Peek N, Brownstein JS, Buckeridge DL. Seven pillars of precision digital health and medicine. Artif Intell Med 2020 Mar;103:101793. [doi: 10.1016/j.artmed.2020.101793] [Medline: 32143798]
33. Olusanya OA, Ammar N, Davis RL, Bednarczyk RA, Shaban-Nejad A. A digital personal health library for enabling precision health promotion to prevent human papilloma virus-associated cancers. Front Digit Health 2021 Jul 21;3:683161 [FREE Full text] [doi: 10.3389/fdgth.2021.683161] [Medline: 34713154]
34. Olusanya OA, White B, Melton CA, Shaban-Nejad A. Examining the implementation of digital health to strengthen the COVID-19 pandemic response and recovery and scale up equitable vaccine access in African countries. JMIR Form Res 2022 May 17;6(5):e34363 [FREE Full text] [doi: 10.2196/34363] [Medline: 35512271]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers

Edited by C Basch; submitted 20.06.22; peer-reviewed by I Kagashe, M Yousef; comments to author 28.07.22; revised version received 18.08.22; accepted 15.09.22; published 17.10.22.

Please cite as:
Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A
Fine-tuned Sentiment Analysis of COVID-19 Vaccine–Related Social Media Data: Comparative Study
J Med Internet Res 2022;24(10):e40408
URL: https://www.jmir.org/2022/10/e40408
doi:10.2196/40408
PMID:36174192

Original Paper

The Potential Role of an Adjunctive Real-Time Locating System in Preventing Secondary Transmission of SARS-CoV-2 in a Hospital Environment: Retrospective Case-Control Study

Min Hyung Kim[1], MD, PhD; Un Hyoung Ryu[2], BS; Seok-Jae Heo[3], PhD; Yong Chan Kim[1], MD, PhD; Yoon Soo Park[1], MD, PhD

[1]Division of Infectious Diseases, Department of Internal Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin-si, Republic of Korea

[2]Division of Planning and Management, Office of Medical Information Technology, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin-si, Republic of Korea

[3]Division of Biostatistics, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Yoon Soo Park, MD, PhD
Division of Infectious Diseases, Department of Internal Medicine
Yongin Severance Hospital
Yonsei University College of Medicine
Dongbaekjukjeon-daero
Giheung-gu
Yongin-si, 16995
Republic of Korea
Phone: 82 10 8502 5825
Email: ysparkok2@yuhs.ac

Abstract

Background: There has been an increasing demand for new technologies regarding infection control in hospital settings to reduce the burden of contact tracing.

Objective: This study aimed to compare the validity of a real-time locating system (RTLS) with that of the conventional contact tracing method for identifying high-risk contact cases associated with the secondary transmission of SARS-CoV-2.

Methods: A retrospective case-control study involving in-hospital contact cases of confirmed COVID-19 patients, who were diagnosed from January 23 to March 25, 2022, was conducted at a university hospital in South Korea. Contact cases were identified using either the conventional method or the RTLS. The primary endpoint of this study was secondary transmission of SARS-CoV-2 among contact cases. Univariate and multivariable logistic regression analysis comparing test positive and versus negative contact cases were performed.

Results: Overall, 509 and 653 cases were confirmed by the conventional method and the RTLS, respectively. Only 74 contact cases were identified by both methods, which could be attributed to the limitations of each method. Sensitivity was higher for the RTLS tracing method (653/1088, 60.0%) than the conventional tracing method (509/1088, 46.8%) considering all contact cases identified by both methods. The secondary transmission rate in the RTLS model was 8.1%, while that in the conventional model was 5.3%. The multivariable logistic regression model revealed that the RTLS was more capable of detecting secondary transmission than the conventional method (adjusted odds ratio 6.15, 95% CI 1.92-28.69; *P*=.007).

Conclusions: This study showed that the RTLS is beneficial when used as an adjunctive approach to the conventional method for contact tracing associated with secondary transmission. However, the RTLS cannot completely replace traditional contact tracing.

(J Med Internet Res 2022;24(10):e41395) doi:10.2196/41395

KEYWORDS

real-time locating system; COVID-19; contact tracing; secondary transmission; SARS-CoV-2

Introduction

Human history is characterized by the incessant influence of infectious diseases, with viruses being the most successful contender [1]. Viral diseases dominate on the World Health Organization's list of top priorities of concern [2]. SARS-CoV-2 is one of the novel viruses exerting unprecedented influence on the world's population due in part to a lack of knowledge. In the early stage of the COVID-19 pandemic, nonpharmaceutical approaches, such as mask wearing and isolating infected patients, comprised the predominant methods of preventing the disease from spreading [3,4]. Despite the development of pharmaceutical agents, such as vaccines and therapeutic antiviral drugs, these nonpharmaceutical measures are not obsolete because of the emergence of new variants and the waning effect of vaccines [5-8]. The importance of nonpharmaceutical measures is emphasized in the hospital environment where immunocompromised populations, such as patients with cancer and older patients, are concentrated.

Contact tracing is an important strategy to keep disease transmission under control by isolating high-risk contacts who eventually present with the disease. Contact tracing is a time- and labor-consuming procedure, the efficiency of which is dependent on the commitment of the infection control personnel and the presence of beneficial adjunctive tools. Moreover, the importance of this method could be diminished in the era of "living with COVID-19," wherein efforts to confine the spread of the disease are diluted with the weakening of disease severity. However, in a hospital setting, simplifying rather than eliminating the contact tracing effort is required. SARS-CoV-2 is a complicated virus to deal with, especially in the hospital setting, without sufficient data regarding its mode of transmission [9-13]. Furthermore, a variety of common transmissible diseases that require contact tracing can spread in hospitals. For most of those diseases, the distance and duration of exposure are of primary importance when deciding the high-risk contacts of a confirmed patient [14-17].

Technological efforts, such as a real-time locating system (RTLS), could be an option to overcome the limitation set by the conventional method. One type of RTLS involves radio-frequency identification (RFID) and a Wi-Fi tracking system. RFID calculates the distance and duration of human-to-human interaction by analyzing the signals from RFID tags worn by users, which are captured by exciters installed in hospital wards and working places [18]. Using this technology, the quantity of interaction affordable, regardless of the number of contacts, can be determined [19]. Evidence of the validity of this technology in a hospital setting is accumulating, despite its privacy concerns and cost-benefit issues [20,21]. The efficiency of this technology for preventing the spread of transmissible diseases needs to be elucidated.

The aim of this study was to evaluate the validity of the RTLS compared with the conventional contact tracing method to identify high-risk contacts associated with the secondary transmission of SARS-CoV-2. Furthermore, we attempted to characterize the factors associated with secondary SARS-CoV-2 transmission using both methods.

Methods

Setting

This study was conducted at a Yongin Severance University–affiliated hospital in South Korea, with 580 beds and an 82% average occupancy rate annually. This institution had RTLS location sensors since its opening in 2020. All health care workers and inpatients were issued RTLS tags that detected their locations.

From the time COVID-19 was declared a pandemic, health care workers, hospitalized patients, and caregivers in this hospital were monitored for the presence of symptoms suggestive of COVID-19 on a daily basis. Caregivers included the patients' family members or privately employed carers. Employees were mandated to report COVID-19–related symptoms through a mobile app at least once a day. Hospitalized patients and caregivers were obligated to take screening reverse transcription polymerase chain reaction (RT-PCR) tests ahead of admission, and COVID-19–related symptoms were closely monitored by attending nurses, which were recorded electronically. SARS-CoV-2 polymerase chain reaction (PCR) tests were conducted for those who developed COVID-19–related symptoms, and quarantine measures were implemented for individuals who tested positive for SARS-CoV-2. Subsequent contact tracing was carried out by the infection control office staff and the digital information team, with stratification of contacts according to the level of exposure. All contacts regardless of the exposure level were recommended to get tested for SARS-CoV-2 at least once, with specific emphasis on high-risk contacts within 14 days after exposure or on the development of symptoms. Postexposure measures, such as quarantine, were implemented for those who were identified as contacts at the discretion of staff in the infection control office. As RTLS data were collected only for research purposes, no postexposure interventions were implemented based solely on RTLS data.

Study Design and Identification of Contact Cases

A retrospective case-control study involving in-hospital contact cases of confirmed COVID-19 patients, who were diagnosed from January 23 to March 25, 2022, was conducted. All contact cases of health care workers, and inpatients and their caregivers, identified either by the conventional method or the RTLS, were included in this study. The participants were followed up from the date of contact to 14 days following the last contact or the date of a follow-up SARS-CoV-2 PCR test.

Contact tracing started 2 days prior to the symptom onset or positive PCR test result of a COVID-19–confirmed patient. The conventional method of contact tracing starts with an in-person interview, followed by reviewing electronic medical records and monitoring closed circuit surveillance camera feeds based on the information acquired in the interview. RTLS-based contact tracing was performed separately by the digital information team. When a patient tested positive, the digital information team extracted data from the RTLS to identify close contact cases. The radio-frequency RTLS sensors that can detect signals within a radius of 20 meters were located in every room in the hospital and at every 10 meters in open spaces. Hospital

staff and inpatients were required to wear RTLS tags at all times. Signals were emitted from the tags every 1 to 3 seconds to confirm the presence of individuals in a room or confirm the distance between individuals through tag-tag signal interaction. When 2 individuals got close enough to a designated distance, the calculation of contact time was started to obtain the cumulative contact time between the 2 individuals. Generally, it took less than 30 minutes to draw data from the RTLS.

The level of exposure was determined based on a Centers for Disease Control and Prevention (CDC) guideline [22]. The CDC provides information on the transmission risk of COVID-19 among contacts according to the level of exposure, and recommends actions to prevent disease transmission. High-risk exposure was defined as close contact with confirmed patients within 2 meters for more than 15 minutes without adequate mask wearing, or physical contact without wearing gloves or protective gowns. Intermediate-risk exposure was defined as contact with confirmed patients within 2 meters for more than 15 minutes with moderate protective equipment. Low-risk exposure was defined as contact with confirmed patients with adequate protective gear, or contact outside the range of high-risk exposure without protective equipment.

The primary endpoint of this study was secondary transmission of SARS-CoV-2 among contact cases. Secondary transmission was assumed when there was a positive conversion of the SARS-CoV-2 PCR test following a negative test result within 14 days of contact. Those without previous test results were included as well, unless they had evidence of other sources of infection, such as being simultaneously diagnosed with index patients, having a known familial transmission, or showing COVID-19–related symptoms.

Exclusion Criteria

The exclusion criteria were as follows: (1) no identifiable age or gender information; (2) no follow-up PCR results; and (3) distance of more than 3 meters from index patients among RTLS-confirmed cases.

Data Collection

The data of in-hospital–confirmed COVID-19 patients were collected retrospectively. Data, including age, gender, vaccination history (including number of vaccinations and days passed from the last vaccination), follow-up SARS-CoV-2 PCR test results, date of diagnosis in case of a positive result, closest exposure distance, duration of exposure within a distance of 2 meters, whether personal protective equipment was used, mask-wearing habit, type of occupation, type of occupation of index patients, date of the last contact, whether the room was shared with index patients, methods used to identify contact cases and postexposure measures, and whether tags were worn, were collected by reviewing the records acquired for contact tracing. We filled up some part of the data regarding wearing masks through estimation based on the hospital policy, when fact-checking was impossible due to long hours of exposure.

SARS-CoV-2 RT-PCR

SARS-CoV-2 RT-PCR was performed with nasopharyngeal swab samples collected from participants. The MagNA Pure 96 System (Roche Diagnostics) was used to extract RNA from nasopharyngeal swabs in viral transport media, according to the manufacturer's instructions. The extracted RNA was then subjected to the Allplex SARS-CoV-2 Assay, which targets 4 genes in a single tube (E, N, RdRp, and S genes) to detect SARS-CoV-2 infection, according to the manufacturer's instructions. PCR amplification was performed using the CFX96 real-time PCR detection system (Bio-Rad Laboratories).

Statistical Methods

Analyses comparing secondary transmission cases and test-negative cases were performed. We allowed the inclusion of multiple episodes of the same individuals because the nature of contact cases was different each time. Baseline characteristics were compared using the Mann-Whitney *U* test, independent samples *t* test, or ANOVA for continuous variables, and the χ^2 test or Fisher exact test for categorical variables. Continuous variables were expressed as means or medians (IQRs) and categorical variables as numbers with percentages for the description of baseline characteristics. Logistic regression was used to identify factors related with secondary transmission adjusting for relevant variables with a *P* value <.05 in univariate analysis. Cumulative hazard curves were created using the Kaplan-Meier method, and the hazards of detecting secondary transmission for each model were compared according to the study date using the log-rank test. Sensitivity analysis was conducted with participants having follow-up PCR results within 14 days available to identify either consistency of or differences in the magnitude of the effect. Missing values were removed from the analysis. All statistical analyses were performed using SPSS software version 26 (IBM Corp). Two-sided *P* values <.05 were considered statistically significant.

Ethics Statement

This study was approved by the Institutional Review Board of Yonsei University Health System Clinical Trial Centre, and the study protocol adhered to the tenets of the Declaration of Helsinki. As the study was retrospective, the Institutional Review Board waived the requirement for written informed consent from the participants (approval number: 9-2022-0027; approved on April 22, 2022).

Results

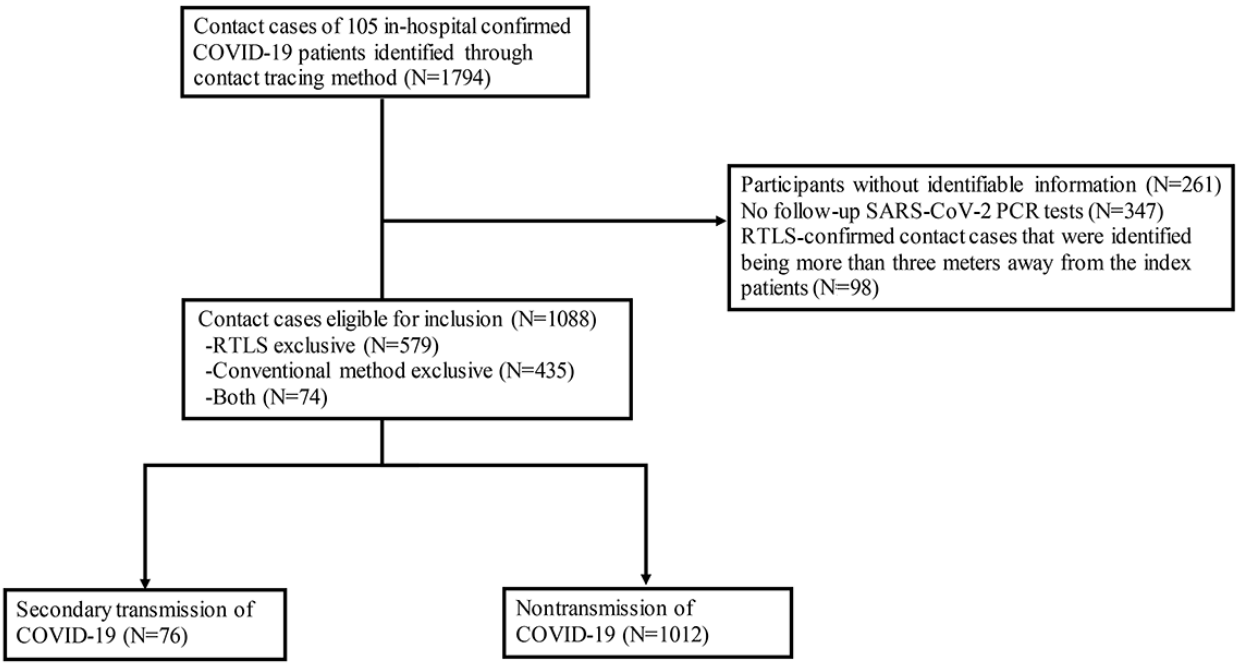
Study Participants

Among 1794 cases identified by the methods described above, 261 cases without age or gender information, 98 RTLS-confirmed cases that were identified more than 3 meters away from index patients, and 347 cases without follow-up test results were excluded. As a result, 1088 contact cases were included in the analysis. Among 79 cases that tested positive for SARS-CoV-2 within 14 days of exposure, 3 cases were excluded from secondary transmission owing to the presence of other sources of transmission (Figure 1).

The baseline characteristics of all participants are presented in the following text. The mean age of the participants was 41.5 (SD 17.5) years, with 25.3% (275/1088) being male participants. Among contacts, 70.7% (769/1088) were health care workers,

and among these, 6.2% (48/769) were doctors, 71.5% (550/769) were nurses, and 22.2% (171/769) were others. The percentage of those vaccinated at least once prior to contact was 83.8% (741/884), with a median of 82 (IQR 54-82) days from the last vaccination. The median contact duration was 240 (IQR 41-1675.8) minutes. Among those who were designated as index patients, 80.3% (843/1051) were health care workers. Furthermore, room sharing was confirmed in 71.9% (736/1023) of cases in the hospital. There were 63.5% (436/686) high-risk contact cases, 28.9% (198/686) intermediate-risk contact cases, and 7.6% (52/686) low-risk contact cases (Multimedia Appendix 1).

Figure 1. Study flow of enrollment. PCR: polymerase chain reaction; RTLS: real-time locating system.



Contact Cases Identified by the RTLS or the Conventional Method

Among the 1088 cases involving 764 participants, 76 cases involving 65 participants resulted in secondary transmission, while 1012 cases involving 730 participants remained negative for SARS-CoV-2 (Figure 1). Only 74 contact cases were identified by both methods, while 509 and 653 cases were confirmed by the conventional and RTLS methods, respectively. The factors associated with RTLS detection against the conventional method were identified. Younger age (37.6, SD 14.5 vs 47.6, SD 19.6 years; *P*<.001), being a health care worker (88.5% vs 60.3%; *P*<.001), being a health care worker who was an index patient (91.4% vs 62.4%; *P*<.001), room sharing (75.5% vs 69.9%; *P*=.04), mask wearing (26.5% vs 41.6%; *P*<.001), and being exposed for long durations (391 [IQR 64-1804] vs 33 [IQR 10-240] minutes; *P*<.001) were associated with RTLS detection. The level of exposure should be interpreted with caution, considering the large number of missing values (Table 1). The absence of records on exposure time largely contributed to the missing values of exposure levels in the conventional method. Analysis involving secondary transmission cases alone revealed that more health care workers were detected by the RTLS exclusive method (Multimedia Appendix 2).

Table 1. Baseline characteristics of participants based on the contact tracing method.

Characteristic	RTLS[a] method[b] (n=653)	Conventional method[b] (n=509)	*P* value
Age (years), mean (SD)	37.6 (14.5)	47.6 (19.6)	<.001
Sex (male), n (%)	164 (25.1)	134 (26.3)	.64
Exposure duration (minutes), median (IQR)	391 (64-1804)	33 (10-240)	<.001
Unknown[c], n	2	344	
Personal protective equipment used			
Mask, n (%)	158 (26.5)	207 (41.6)	<.001
Gloves, n (%)	0 (0.0)	6 (3.6)	.59
Face shield, n (%)	0 (0.0)	1 (0.6)	>.99
Unknown[c], n	56	11	
Mask-wearing consistency[d]			
At all times, n (%)	23 (92.0)	138 (84.7)	.33
More than 50%, n (%)	1 (4.0)	10 (6.1)	>.99
Less than 50%, n (%)	1 (4.0)	15 (9.2)	.70
Unknown[c], n	628	346	
Level of exposure			
High, n (%)	422 (70.9)	54 (33.3)	<.001
Intermediate, n (%)	161 (27.1)	61 (37.7)	<.001
Low, n (%)	12 (2.0)	47 (29.0)	<.001
Unknown[c], n	58	347	
Type of occupation			<.001
Health care worker, n (%)	525 (88.5)	284 (60.3)	
Doctor, n (%)	26 (4.4)	23 (4.9)	
Nurse, n (%)	388 (65.4)	188 (39.9)	
Patient, n (%)	68 (11.5)	187 (39.7)	
Patient, n (%)	68 (11.5)	130 (27.6)	
Caregiver, n (%)	0 (0.0)	57 (12.1)	
Unknown[c], n	60	38	
Type of occupation of the index patient			<0.001
Health care worker, n (%)	597 (91.4)	294(62.4)	
Patient, n (%)	56 (8.6)	177 (37.6)	
Unknown[c], n	0	38	
Vaccination status			
Vaccinated more than once, n (%)	493 (81.9)	294 (86.5)	.07
Days from last vaccination[e] (days), median (IQR)	83 (68-196)	60 (38-170)	<.001
Unknown[c], n	51	169	
Postexposure measure			
Quarantined, n (%)	23 (69.7)	127 (65.8)	.06
Monitored actively, n (%)	8 (24.2)	34 (17.6)	.31
Monitored passively, n (%)	2 (6.1)	32 (16.6)	.12
Unknown[c], n	620	346	

Characteristic	RTLS[a] method[b] (n=653)	Conventional method[b] (n=509)	*P* value
Secondary transmission, n (%)	53 (8.1)	31 (6.1)	.06
Room sharing, n (%)	450 (75.5)	348 (69.9)	.04
Unknown[c], n	57	11	
Compliance with tag wearing, n (%)	653 (100.0)	246 (48.3)	<.001

[a]RTLS: real-time locating system.

[b]Cases included in both the RTLS and conventional methods were handled as duplicate values.

[c]Unknown represents the number of missing values.

[d]Extent to which each participant conforms to the mask-wearing precaution.

[e] Days passed from the last vaccination.

Secondary Transmission Among Contact Cases Identified by the RTLS or the Conventional Method

The baseline characteristics of secondary transmission cases are described in Multimedia Appendix 1. Overall, the secondary transmission rate was 7.0% when all contact tracing methods were combined. The secondary transmission rate in the RTLS model was 8.1%, while that in the conventional method model was 5.3% (Table 2). The results spread out according to the confirmed date are presented in Figure 2, which shows a higher contribution of the RTLS than the conventional method in detecting secondary transmission.

We calculated the odds ratio (OR) for the secondary transmission group, with the group that tested negative for SARS-CoV-2 as a control. Variables with clinical significance and statistical significance in the univariate analysis were included in the multivariate analysis. The adjusted odds ratios (aORs) for clinically relevant variables and for variables with statistical significance in the univariate analysis revealed that male gender (aOR 0.11, 95% CI 0.01-0.53; *P*=.03), longer duration from the last vaccination (aOR 1.04, 95% CI 1.01-1.07; *P*=.006), and using the RTLS as the contact tracing method (aOR 6.15, 95% CI 1.92-28.69; *P*=.007) were associated with secondary transmission (Table 3). The Kaplan-Meier curve showed increased detection of secondary transmission among contact cases identified by the RTLS toward the end of the study period (Multimedia Appendix 3). Moreover, a subgroup analysis involving contact cases with available follow-up PCR tests within 14 days produced similar results (Multimedia Appendix 4).

The difference in cumulative contact duration was not statistically significant between the groups. The median contact duration was 630 [IQR 72.5-1510.5] minutes for the cases having secondary transmission versus 240 [IQR 41-1678] minutes for the controls (Multimedia Appendix 1). There were 3 cases of secondary transmission with 15 minutes of contact within 2 meters. All 3 were identified by the RTLS, and the time was precisely calculated. None of the cases were involved in aerosol-producing procedures.

Table 2. Comparison of the performance of each contact tracing method and the methods combined.

Variable	RTLS[a] method		Conventional method		Both methods[b]	
	Detected	Not detected	Detected	Not detected	Detected	Not detected
Identified contact cases (N=1088), n (%)	653 (60.0)	435 (40.0)	509 (46.8)	579 (53.2)	74 (6.8)	1,014 (93.2)
Secondary transmission (N=76), n (%)	53 (69.7)	23 (30.3)	27 (35.5)	49 (64.5)	4 (5.3)	72 (94.7)
Secondary transmission rate[c], %	8.1	N/A[d]	5.3	N/A	5.4	N/A

[a]RTLS: real-time locating system.

[b]"Both methods" denotes cases identified by both the RTLS and conventional methods.

[c]Secondary transmission rate was defined as cases of secondary transmission against contact cases identified by each method.

[d]N/A: not applicable.

Figure 2. Secondary transmission rate calculated against contacts identified by each method according to the date of diagnosis. The secondary transmission rate was defined as cases of secondary transmission against contacts according to the date of index patients' confirmation. The average secondary transmission rate calculated against contacts identified exclusively by the RTLS was 10.6%, while that calculated against contacts identified exclusively by the conventional method was 7%. "Both" denotes cases identified by both the RTLS and conventional method. RTLS: real-time locating system.

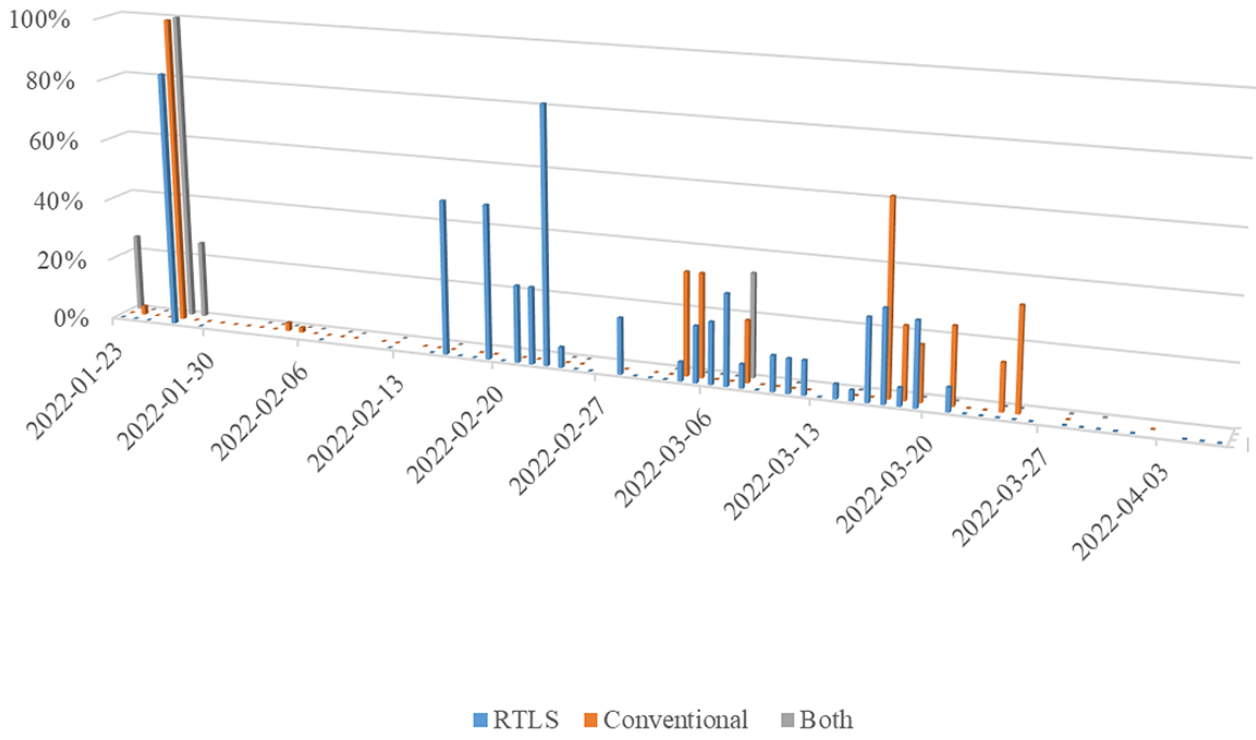


Table 3. Logistic regression analysis for identifying the risk factors for SARS-CoV-2 secondary transmission.

Variable	Univariate analysis			Multivariable analysis[a]		
	OR[b]	95% CI	*P* value	OR[c]	95% CI	*P* value
Age	0.97	0.93-1.00	.12	1.00	0.96-1.05	.90
Male (reference: female)	0.11	0.01-0.50	.03	0.11	0.01-0.53	.03
Days from the last vaccination[d] (days)	1.05	1.02-1.07	.001	1.04	1.01-1.07	.006
Room sharing	1.96	0.85-5.32	.14	2.72	0.40-14.42	.26
Mask wearing	0.55	0.23-1.19	.15	2.20	0.35-9.86	.34
RTLS[e] (reference: conventional)	5.94	2.09-24.92	.004	6.15	1.92-28.69	.007

[a]Logistic regression was used to calculate the risk of secondary transmission. Variables with clinical significance and statistical significance in the univariate analysis were included in the model.

[b]OR: odds ratio.

[c]Adjustment for all the variables involved in the univariate model.

[d]Days passed from the last vaccination.

[e]RTLS: real-time locating system.

Discussion

Principal Findings

This study suggested that the RTLS has an added benefit for identifying close contact cases associated with secondary transmission of SARS-CoV-2 by identifying 64.5% (49/76) additional cases that were not detected by the conventional method. The RTLS had a higher power than the conventional method for detecting high-risk contact cases that subsequently developed COVID-19. However, the technology may not be used separately from the conventional method owing to moderate sensitivity.

Comparison With Prior Work

The utility of the RTLS for tracing contacts of multiple transmissible diseases in a hospital setting has been explored in previous studies. Researchers suggested that the RTLS has

moderate to high sensitivity but a low positive predictive value when compared with the conventional tracing method in detecting contact cases for droplet-transmitted diseases such as COVID-19 [21,23]. Our study showed that sensitivity was higher for the RTLS tracing method (653/1088, 60.0%) than the conventional tracing method (509/1088, 46.8%) considering all contact cases identified by both methods. The value is not acceptably high for its use as a single method for contact tracing. However, this method showed promising results in terms of efficiency. To the best of our knowledge, this study is the first to discuss the efficiency of the RTLS for detecting high-risk contact cases associated with secondary SARS-CoV-2 infection. When all methods were combined, the secondary transmission rate among health care staff and patients was 7.0%, while that identified by the RTLS was 8.1% and that identified by the conventional method was 5.3%. The average secondary transmission rate was lower than that in the community setting [24-26]. Based on the fact that the denominator involves contact cases identified by methods with no known gold standard, a higher secondary transmission rate may mean higher efficiency of the contact tracing method. The logistic regression model showed that the odds of detecting secondary transmission cases was higher for the RTLS than the conventional method. This may indicate that the RTLS is not inferior to conventional methods in performing contact tracing, especially when considering its time-saving characteristics.

The sensitivity of the RTLS was lower when the conventional method was used as a reference (74/653, 11.3%). This discrepancy is associated with inherent limitations of the RTLS or conventional method. The efficacy of the RTLS is dependent on the commitment of participants to wearing tags and the frequency of the signal exciter [27]. As was shown in this study, the tag-wearing behavior and location of participants were associated with the discrepancy. Working as a young nurse was associated with RTLS detection owing to a favorable tag-wearing behavior. Conventional contact tracing relies heavily on a person's memory, which might be subjective and inaccurate. It tends to be biased toward identifying vulnerable contacts, such as hospitalized patients, which may be another explanation for the discrepancy. The RTLS would be beneficial when used for highly transmissible infectious diseases because of its time-saving property, which can help detect more high-risk contacts associated with secondary transmission. The Kaplan-Meier curve showed a trend of increased detection of secondary transmission cases through the RTLS toward the end of our research when an increasing proportion of Omicron variant cases was being reported on a weekly basis. However, our results suggest that the RTLS cannot be used alone for tracing contacts. Although the efficacy of the RTLS as an adjunctive approach to the conventional method was noted, separate and solitary use of the RTLS has not been verified. The fact that nearly 40.0% (435/1088) of contact cases and 30.3% (23/76) of secondary contact cases could have been missed without the conventional method is worth noting. Based on the results of the analysis (Table 1) and the Kaplan-Meier curve (Multimedia Appendix 3), we recommend using the RTLS when tracing the contacts of persons with highly contagious diseases, who are likely to wear tags, such as nursing staff, and who share the same space for a long time.

Factors indicative of prolonged exposure to the symptomatic source, such as room sharing and mask-wearing behavior, were not associated with secondary transmission, which was inconsistent with the findings of previous studies [28,29]. On the other hand, being female increased the risk of secondary transmission. This may be because nursing staff members were mostly women at the institution and were involved in activities that had high risks of transmission. Detailed information should be collected to discuss the risk of transmission.

It needs to be noted that contact duration was not statistically different between the 2 groups. This study has advantages in determining the significance of contact duration in the transmission of the disease owing to the implementation of methods capable of quantifying time precisely. The average time spent with confirmed patients was long, which is plausible, considering the interactions taking place between individuals in health care facilities. Maintaining strict precautions, such as frequent hand washing, would be crucial for preventing disease spread when the cumulative time surpasses a certain extent, taking into consideration previous studies that emphasized the role of fomites in disease transmission [30,31]. Furthermore, there were 3 cases of transmission with less than 15 minutes of contact time, which has been designated as a transmission cutoff by the CDC. Considering that the number of participants working in the high-risk department was not significantly different (data not shown), aerosol-producing procedures were not attributable to the finding, even though there was a risk of transmission owing to high-risk behaviors, such as coughing and sneezing. In light of previous reports indicating the failure of containment of the disease with existing guidelines [32], further efforts to elucidate the threshold of the transmission of SARS-CoV-2 are warranted. The RTLS could be used for research purposes to better characterize the transmission rate of a novel disease or variant, thereby guiding institutional and government policies.

Limitations

This study has some limitations that must be acknowledged. First, the assumption of index cases may not be completely accurate without a genetic analysis [25,29], especially considering the high incidence of COVID-19 cases in the community. Second, owing to the retrospective design of the study, we could not confirm the extent of use of personal protective equipment and the presence of symptoms, especially for the contact cases identified by the RTLS. Third, we could not accurately calculate the positive predictive value of the RTLS contact tracing model because of the lack of verification. Finally, we should take into consideration the cost of installation of the RTLS, which may not be feasible in a resource-limited setting. However, this study is significant in that it investigated the utility of a novel technology in contact tracing in the backdrop of a hospital environment, reflecting the real-world circumstances where disease transmission actually takes place. Our findings underscore the need for further studies investigating the efficiency of the technology with prospectively collected data.

Conclusions

This study showed that novel technologies, such as the RTLS, are beneficial when used as an adjunctive approach to the conventional method for contact tracing, especially when individuals share rooms with each other and under the influence of highly transmissible diseases. However, the RTLS cannot completely replace the traditional contact tracing method.

Acknowledgments

We would like to thank all of the nursing and infection control office staff, as well as the physicians who supported this project. Finally, we give credit to all of the patients who took part in this study. This study was supported by a faculty research grant of Yonsei University College of Medicine (6-2022-0088).

Authors' Contributions

MHK was responsible for data analysis and contributed to the drafting and writing of the manuscript. YSP was the chief investigator and was responsible for the conception and design of the study. UHR and SJH were involved in data acquisition and interpretation. YCK reviewed the data and provided feedback. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics of participants with and without secondary SARS-CoV-2 transmission.
[DOCX File , 24 KB - jmir_v24i10e41395_app1.docx]

Multimedia Appendix 2
Comparison of contact tracing methods among secondary transmission cases.
[DOCX File , 21 KB - jmir_v24i10e41395_app2.docx]

Multimedia Appendix 3
Kaplan-Meier curve for the risk of secondary transmission according to the tracing method based on the study date. The cumulative hazard of secondary transmission is shown for the RTLS method vs the conventional method (P=.001). RTLS: real-time locating system.
[PNG File , 68 KB - jmir_v24i10e41395_app3.png]

Multimedia Appendix 4
Subgroup analysis involving contact cases that had follow-up SARS-CoV-2 polymerase chain reaction results within 14 days.
[DOCX File , 19 KB - jmir_v24i10e41395_app4.docx]

References

1. Piret J, Boivin G. Pandemics throughout history. Front Microbiol 2020 Jan 15;11:631736 [FREE Full text] [doi: 10.3389/fmicb.2020.631736] [Medline: 33584597]
2. Prioritizing diseases for research and development in emergency contexts. World Health Organization. URL: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts [accessed 2022-10-07]
3. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug 08;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
4. Chowdhury R, Heng K, Shawon MSR, Goh G, Okonofua D, Ochoa-Rosales C, Global Dynamic Interventions Strategies for COVID-19 Collaborative Group. Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. Eur J Epidemiol 2020 May 19;35(5):389-399 [FREE Full text] [doi: 10.1007/s10654-020-00649-w] [Medline: 32430840]
5. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. N Engl J Med 2022 Apr 21;386(16):1532-1546 [FREE Full text] [doi: 10.1056/NEJMoa2119451] [Medline: 35249272]
6. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis. The Lancet Infectious Diseases 2022 May;22(5):603-610. [doi: 10.1016/s1473-3099(22)00001-9]
7. Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Amir O, Freedman L, et al. Protection by a fourth dose of BNT162b2 against Omicron in Israel. N Engl J Med 2022 May 05;386(18):1712-1720. [doi: 10.1056/nejmoa2201570]

8. Regev-Yochay G, Gonen T, Gilboa M, Mandelboim M, Indenbaum V, Amit S, et al. Efficacy of a fourth dose of Covid-19 mRNA vaccine against Omicron. N Engl J Med 2022 Apr 07;386(14):1377-1380. [doi: 10.1056/nejmc2202542]
9. Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. Proc Natl Acad Sci U S A 2020 Jun 30;117(26):14857-14863 [FREE Full text] [doi: 10.1073/pnas.2009637117] [Medline: 32527856]
10. Morawska L, Milton D. It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). Clin Infect Dis 2020 Dec 03;71(9):2311-2313 [FREE Full text] [doi: 10.1093/cid/ciaa939] [Medline: 32628269]
11. Drahl C. A conversation with Jose-Luis Jimenez. ACS Cent Sci 2020 Dec 23;6(12):2118-2119 [FREE Full text] [doi: 10.1021/acscentsci.0c01575] [Medline: 33376771]
12. Wilson N, Corbett S, Tovey E. Airborne transmission of covid-19. BMJ 2020 Aug 20;370:m3206. [doi: 10.1136/bmj.m3206] [Medline: 32819961]
13. de Man P, Paltansing S, Ong D, Vaessen N, van Nielen G, Koeleman J. Outbreak of coronavirus disease 2019 (COVID-19) in a nursing home associated with aerosol transmission as a result of inadequate ventilation. Clin Infect Dis 2021 Jul 01;73(1):170-171 [FREE Full text] [doi: 10.1093/cid/ciaa1270] [Medline: 32857130]
14. Glennie M, Gardner K, Dowden M, Currie BJ. Active case detection methods for crusted scabies and leprosy: A systematic review. PLoS Negl Trop Dis 2021 Jul 23;15(7):e0009577 [FREE Full text] [doi: 10.1371/journal.pntd.0009577] [Medline: 34297724]
15. Begun M, Newall AT, Marks GB, Wood JG. Contact tracing of tuberculosis: a systematic review of transmission modelling studies. PLoS One 2013 Sep 4;8(9):e72470 [FREE Full text] [doi: 10.1371/journal.pone.0072470] [Medline: 24023742]
16. Otter J, Mookerjee S, Davies F, Bolt F, Dyakova E, Shersing Y, et al. Detecting carbapenemase-producing Enterobacterales (CPE): an evaluation of an enhanced CPE infection control and screening programme in acute care. J Antimicrob Chemother 2020 Sep 01;75(9):2670-2676. [doi: 10.1093/jac/dkaa192] [Medline: 32479615]
17. Transmission of SARS-CoV-2: implications for infection prevention precautions. World Health Organization. URL: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions [accessed 2022-10-07]
18. Camacho-Cogollo JE, Bonet I, Iadanza E. RFID technology in health care. In: Clinical Engineering Handbook. San Diego, CA: Elsevier; 2020:33-41.
19. Bowen ME, Wingrave CA, Klanchar A, Craighead J. Tracking technology: Lessons learned in two health care sites. THC 2013 Jun 01;21(3):191-197. [doi: 10.3233/thc-130738]
20. Patel B, Vilendrer S, Kling SMR, Brown I, Ribeira R, Eisenberg M, et al. Using a real-time locating system to evaluate the impact of telemedicine in an emergency department during COVID-19: Observational study. J Med Internet Res 2021 Jul 26;23(7):e29240 [FREE Full text] [doi: 10.2196/29240] [Medline: 34236993]
21. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim W, Chow A. Use of a real-time locating system for contact tracing of health care workers during the COVID-19 pandemic at an infectious disease center in Singapore: Validation study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
22. Interim Infection Prevention and Control Recommendations for Healthcare Personnel During the Coronavirus Disease 2019 (COVID-19) Pandemic. CDC. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html [accessed 2022-10-07]
23. Huang Z, Guo H, Lee Y, Ho EC, Ang H, Chow A. Performance of digital contact tracing tools for COVID-19 response in Singapore: Cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e23148 [FREE Full text] [doi: 10.2196/23148] [Medline: 33006944]
24. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: An updated systematic review and meta-analysis. JAMA Netw Open 2022 Apr 01;5(4):e229317 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.9317] [Medline: 35482308]
25. Czech-Sioli M, Günther T, Robitaille A, Roggenkamp H, Büttner H, Indenbirken D, et al. Integration of sequencing and epidemiological data for surveillance of SARS-CoV-2 infections in a tertiary-care hospital. Clin Infect Dis 2022 Jun 19:2022 [FREE Full text] [doi: 10.1093/cid/ciac484] [Medline: 35717654]
26. Sikkens JJ, Buis DTP, Peters EJG, Dekker M, Schinkel M, Reijnders TDY, et al. Serologic surveillance and phylogenetic analysis of SARS-CoV-2 infection among hospital health care workers. JAMA Netw Open 2021 Jul 01;4(7):e2118554 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18554] [Medline: 34319354]
27. Anglemyer A, Moore T, Parker L, Chambers T, Grady A, Chiu K, et al. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst Rev 2020 Aug 18;8:CD013699 [FREE Full text] [doi: 10.1002/14651858.CD013699] [Medline: 33502000]
28. Kim D, Lin M, Jennings C, Li H, Jung J, Moore N, CDC Prevention Epicenter Program. Duration of replication-competent SARS-CoV-2 shedding among patients with severe or critical coronavirus disease 2019 (COVID-19). Clin Infect Dis 2022 May 24:2022 [FREE Full text] [doi: 10.1093/cid/ciac405] [Medline: 35607802]
29. Chow K, Aslam A, McClure T, Singh J, Burns J, McMillen T, et al. Risk of healthcare-associated transmission of sever acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in hospitalized cancer patients. Clin Infect Dis 2022 May 03;74(9):1579-1585 [FREE Full text] [doi: 10.1093/cid/ciab670] [Medline: 34329418]

30. Ren C, Xi C, Wang J, Feng Z, Nasiri F, Cao S, et al. Mitigating COVID-19 infection disease transmission in indoor environment using physical barriers. Sustain Cities Soc 2021 Nov;74:103175 [FREE Full text] [doi: 10.1016/j.scs.2021.103175] [Medline: 34306996]
31. Klompas M, Baker M, Rhee C. Coronavirus disease 2019's challenges to infection control dogma regarding respiratory virus transmission. Clin Infect Dis 2022 Aug 24;75(1):e102-e104. [doi: 10.1093/cid/ciac204] [Medline: 35271714]
32. Jung J, Lee J, Kim E, Namgung S, Kim Y, Yun M, et al. Frequent occurrence of SARS-CoV-2 transmission among non-close contacts exposed to COVID-19 patients. J Korean Med Sci 2021 Aug 23;36(33):e233 [FREE Full text] [doi: 10.3346/jkms.2021.36.e233] [Medline: 34427062]

Abbreviations

aOR: adjusted odds ratio
CDC: Centers for Disease Control and Prevention
PCR: polymerase chain reaction
RFID: radio-frequency identification
RTLS: real-time locating system
RT-PCR: reverse transcription polymerase chain reaction

Edited by C Basch; submitted 25.07.22; peer-reviewed by S Vilendrer, A Chow, MS Aslam; comments to author 22.08.22; revised version received 09.09.22; accepted 26.09.22; published 18.10.22.

Please cite as:
Kim MH, Ryu UH, Heo SJ, Kim YC, Park YS
The Potential Role of an Adjunctive Real-Time Locating System in Preventing Secondary Transmission of SARS-CoV-2 in a Hospital Environment: Retrospective Case-Control Study
J Med Internet Res 2022;24(10):e41395
URL: https://www.jmir.org/2022/10/e41395
doi:10.2196/41395
PMID:36197844

Original Paper

Patient and Provider Experiences With Virtual Care in a Large, Ambulatory Care Hospital in Ontario, Canada During the COVID-19 Pandemic: Observational Study

Cherry Chu[1], MSc; Dhruv Nayyar[2,3], MD, MSc; Onil Bhattacharyya[1,2,4,5], MD, PhD; Danielle Martin[1,2,3,4,5], MD, MPP; Payal Agarwal[2,4,5], MD; Geetha Mukerji[1,2,3,4], MD, MSc

[1]Women's College Hospital Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[4]Women's College Hospital, Toronto, ON, Canada

[5]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Geetha Mukerji, MD, MSc
Women's College Hospital Institute for Health System Solutions and Virtual Care
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400
Email: geetha.mukerji@wchospital.ca

Abstract

Background: Virtual care use increased during the COVID-19 pandemic. The impact of that shift on patient and provider experiences is unclear.

Objective: We evaluated patient and provider experiences with virtual visits across an academic, ambulatory hospital in Toronto, Canada and assessed predictors of positive experience with virtual care.

Methods: Survey data were analyzed from consenting patients who attended at least one virtual visit (video or telephone) and from consenting providers who delivered at least one virtual visit. Distributions for demographic variables and responses to survey questions are reported, with statistical significance assessed using chi-square tests and t tests. Ordinal logistic regression analysis was used to identify any patient predictors of responses.

Results: During the study period, 253 patients (mean age 45.1, SD 15.6 years) completed 517 video visit surveys, and 147 patients (mean age 41.6, SD 16.4 years) completed 209 telephone visit surveys. A total of 75 and 94 providers completed the survey in June 2020 and June 2021, respectively. On a scale from 1 to 10 regarding likelihood to recommend virtual care to others, fewer providers rated a score of 8 or above compared with patients (providers: 62/94, 66% for video and 49/94, 52% for telephone; patients: 415/517, 80% for video and 150/209, 72% for telephone). Patients of non-White ethnicity had lower odds of rating a high score of 9 or 10 compared with White patients (odds ratio 0.52, 95% CI 0.28-0.99).

Conclusions: Patient experiences with virtual care were generally positive, but provider experiences were less so. Findings suggest potential differences in patient experience by ethnicity, warranting further investigation into equity concerns with virtual care.

(J Med Internet Res 2022;24(10):e38604) doi:10.2196/38604

KEYWORDS

virtual care; telehealth; COVID-19 pandemic; patient experience; provider experience; virtual; telemedicine; COVID-19; ethnicity; social factors; experience

Introduction

The COVID-19 pandemic has accelerated virtual care use in many jurisdictions, stemming from the need for physical distancing, preservation of personal protective equipment, and the desire to adhere to public health guidance [1-3]. In the province of Ontario, Canada, virtual care adoption was low prior to the COVID-19 pandemic, in large part due to restrictive reimbursement policies [4]. The onset of the pandemic led to the introduction of temporary billing codes allowing for reimbursement of various modalities of virtual care, including videoconferencing across a wide range of platforms as well as telephone visits [5]. During this global crisis, virtual care use in Ontario increased significantly, from 1.6% of total ambulatory visits in the second quarter of 2019 to 70.6% in the second quarter of 2020 [6].

Many studies have been published on patient or provider experiences with virtual care before and during the pandemic [7-12]. Some studies have found positive experiences with virtual visits due to reasons such as convenience and travel time avoided [13-15], while other studies have reported that patients and providers did not find the quality of virtual visits to be better than in-person visits [16]. However, to our knowledge, the literature on patient and provider experiences with virtual care has been mostly limited to small-scale studies localized to a specific clinical program. Furthermore, few have attempted to address potential equity considerations that might contribute to differences in patient experience [17,18]. As virtual care becomes more prevalent, so does the potential issue of the digital divide, in which patients of marginalized populations, such as older age, lower health literacy, non-White ethnic backgrounds, or lower income, may have worse access to or experiences with virtual health services compared with others for reasons such as lack of access to technology or resources in general, discrimination, and limited digital health literacy [19,20].

At Women's College Hospital (WCH), an academic ambulatory hospital in Toronto, Ontario, Canada, virtual care adoption accelerated during the COVID-19 pandemic. In-person ambulatory visits throughout the hospital were largely replaced with video or telephone appointments. In this study, our objective was to use data from WCH during the pandemic to describe patient and provider experiences with virtual care across various clinical areas and to identify demographic characteristics associated with patient experience.

Methods

Context

WCH is an ambulatory, academic facility in Toronto, Canada and is fully affiliated with the University of Toronto. From an administrative perspective, virtual visits are scheduled almost identically to in-person visits, the only difference being the visit type used. For video visits specifically, the electronic medical record is able to automatically create video encounters with the use of an existing platform (ie, Zoom, a video conferencing service that can be licensed for secure use for health care purposes) when specific video visit types are used. When booking a phone visit, it is clearly identified to clinicians that the visit is to take place by phone. The hospital provided resources on its website, notifying patients of the option of video virtual care and its availability as well as how to use it, including training videos and guides. Otherwise, patients would have been presented with the option during phone calls with administrative staff or notified of a virtual video or phone visit within the appointment notification letter they received. Clinicians were encouraged to curate their environment prior to conducting virtual visits, especially by video. In many instances, clinicians continued to work within the clinic when conducting virtual care, though several clinics or departments worked mostly or entirely from home. Regardless, creating private space and ensuring the use of appropriate technology were encouraged. Clinicians were also encouraged to collect appropriate identifying information from patients, much as would occur during check-in during an in-person visit. On the patient end, the resources provided to them motivated them to treat virtual encounters much as they would in-person encounters, including ensuring their own private space, being in a well-lit area, and minimizing background distractions.

Ethical Approval

This study received ethics approval from the WCH Research Ethics Board (REB # 2019-0191-E).

Data Sources and Population

Data were collected from surveys administered to consenting patients who attended at least one virtual visit (video or telephone) at any clinic within WCH who consented via digital consent to be sent a patient experience survey after their visit. Patients had to be registered in the patient portal to receive the digital consent via the electronic medical record. Patients were offered the opportunity to complete the survey after every virtual visit attended. We analyzed patient experience survey data for video visits from May 2020 to May 2021 (253 patients and 517 responses). As survey deployment for telephone visits was delayed due to staff shortages in the second pandemic wave, we analyzed data for telephone visits since survey launch from October 2020 to May 2021 (147 patients and 209 responses). Provider experience surveys were administered twice, in June 2020 and June 2021, for those who delivered at least one virtual visit at any clinic within the hospital. Both patient and provider surveys included questions regarding demographic characteristics and satisfaction with virtual care, with opportunities for written feedback. Patient and provider survey questions can be found in Multimedia Appendix 1.

Statistical Analysis

Descriptive statistics were used to summarize demographic characteristics and responses to patient experience questions among all patients and new patients (defined as those who had an initial visit with a clinic during the study period) who attended at least one virtual visit at the hospital. The following demographic characteristics were self-identified through survey responses: age, gender, ethnicity, household income, and English-language proficiency. Overall marginalization was determined using the Ontario marginalization index (ON-Marg) [21], which was linked to patient postal codes. ON-Marg is a tool that measures deprivation on multiple levels, including

economic, ethnoracial, age-based, and social marginalization. Chi-square tests, Fischer exact tests, and *t* tests were performed as applicable to compare demographic characteristics and responses to various questions regarding patient experience (eg, helpfulness of the virtual visit, likelihood of recommending virtual visits to a friend) between patients who attended video appointments and patients who attended telephone appointments.

The net promoter score, a metric used to measure a client's willingness to recommend a company's product or services [22], was also calculated for video and telephone visits. As the net promoter score can range from –100% to +100%, any score greater than 0% can be considered a desirable score [23]. Ordinal logistic regression analysis was performed on all survey responses to identify any patient demographic variables that may predict their response to select questions on likelihood to recommend virtual visits and perceived helpfulness of the virtual visits. Findings from the regression model compare the odds of choosing an answer in the highest category compared with the other 2 categories (eg, rating of 9-10 compared with ratings of 1-6 and 7-8). Only patients with complete responses to all relevant survey questions were included in the models. Provider ratings to various survey questions on their experience, including the perceived quality of the virtual visit compared with an in-person visit, the amount of time and effort required to conduct the virtual visit, and others, were compared descriptively between June 2020 and June 2021.

Results

Patient Experience

Among all virtual visits for patients registered to the patient portal during the study period, the proportion of individuals who consented to be sent a survey was 1057 of 1872 (56.5%) for video visits and 259 of 358 (72.3%) for phone visits. Among those who consented, 517 of 1057 (48.9%) video visit surveys and 209 of 259 (80.7%) phone visit surveys were completed. Baseline characteristics of all patients who responded to at least one video or telephone survey are reported in Table 1. A total of 253 unique patients completed 517 video surveys, while 147 unique patients completed 209 telephone surveys. There were more women (130/147, 88.4% vs 195/253, 77.1%; *P*=.005) and older patients (mean age 45.1, SD 15.6 years vs mean 41.6, SD 16.4 years; *P*=.04) who completed telephone visit surveys compared with video visit surveys. Among survey respondents who were new patients, telephone users were also older than video users (mean age 50.5, SD 17.5 years vs mean 41.3, SD 15.3 years; *P*=.02; Table S1 in Multimedia Appendix 2). The top clinical departments through which patients attended the virtual visits are listed in Tables S2 and S3 in Multimedia Appendix 2.

Table 2 reports patient responses regarding their experience with virtual care. Most patients found their virtual visit to be very helpful for their health issue (417/517, 80.7% for video and 154/209, 73.7% for telephone). When asked what they would have done if a virtual visit with the doctor was not available, most selected "I would not have sought care at that time" (198/517, 38.3% of video users and 43/209, 20.6% of telephone users), "Scheduled an in-person visit with this doctor" (192/517, 37.1% of video users and 99/209, 47.4% of telephone users), or "See/talk to my family doctor" (145/517, 28.1% of video users and 58/209, 27.8% of telephone users). On a scale of 1 to 10 regarding likelihood to recommend virtual care to a friend, most patients responded with a rating of 8 or above (415/517, 80.3% of video users and 150/209, 71.8% of telephone users). The net promoter score for video visits was higher than that for telephone visits (60.2% vs 40.4%). Findings were similar when only considering new patients (Table S4 in Multimedia Appendix 2). However, the difference in net promoter scores between video and telephone visits was greater for new patients (51.8% for video vs 15.0% for telephone). The majority of patients (405/517, 78.3% for video and 146/209, 69.9% for telephone) preferred to have the option of virtual visits after COVID-19.

We report results for the ordinal logistic regression models in Table 3. In model 1, 255 patients with complete responses were included, and 268 patients were included in model 2. From model 1, ethnicity and age group were significant predictors of the likelihood of recommending video or phone visits to a friend. Specifically, patients who were non-White had lower odds of rating a high score of 9 or 10 compared with White patients (odds ratio [OR] 0.52, 95% CI 0.28-0.99), and patients aged 50 years to 59 years had lower odds of rating a high score of 9 or 10 compared with patients aged 30 years to 39 years (OR 0.26, 95% CI 0.11-0.64). None of the independent variables (ethnicity, family income, overall marginalization, gender, and age) assessed in model 2 appeared to be significant predictors of how helpful the virtual visit was in addressing the patient's health issue.

Table 1. Baseline patient characteristics of all video and phone survey respondents.

Characteristic	Video visit survey respondents (n=253)[a]	Video visit survey respondents among nonmissing respondents, n (%)	Phone visit survey respondents (n=147)[b]	Phone visit survey respondents among nonmissing respondents, n (%)	*P* value[c]
Gender, n (%)					
Female	195 (77.1)	195 (77.1)	130 (88.4)	130 (88.4)	.005
Male	58 (22.9)	58 (22.9)	17 (11.6)	17 (11.6)	
Age (years), mean (SD)	41.6 (16.4)	N/A[d]	45.1 (15.6)	N/A	.04
Total family income (CAD) in previous year[e], n (%)					
0 to 29,999	9 (3.5)	9 (8.3)[f]	8 (5.4)	8 (5.5)[g]	.82
30,000 to 59,999	10 (4.0)	10 (9.3)[f]	12 (8.2)	12 (8.2)[g]	
60,000 to 89,999	12 (4.7)	12 (11.1)[f]	20 (13.6)	20 (13.7)[g]	
90,000 to 119,000	15 (5.9)	15 (13.9)[f]	17 (11.6)	17 (11.6)[g]	
120,000 to 149,000	4 (1.6)	4 (3.7)[f]	10 (6.8)	10 (6.8)[g]	
150,000 or more	26 (10.3)	26 (24.1)[f]	27 (18.4)	27 (18.5)[g]	
Do not know	5 (2.0)	5 (4.6)[f]	8 (5.4)	8 (5.5)[g]	
Prefer not to answer	27 (10.7)	27 (25.0)[f]	44 (29.9)	44 (30.1)[g]	
Missing	145 (57.3)	N/A	1 (0.7)	N/A	
Ethnicity[h], n (%)					
White	82 (32.4)	82 (73.2)[i]	102 (69.4)	102 (70.8)[j]	.54
Asian	9 (3.6)	9 (8.0)[i]	19 (12.9)	19 (13.2)[j]	
Black	4 (1.6)	4(3.6)[i]	4 (2.7)	4 (2.8)[j]	
Latin American	5 (2.0)	5 (4.5)[i]	2 (1.4)	2 (1.4)[j]	
Indigenous	1 (0.4)	1 (0.9)[i]	1 (0.7)	1 (0.7)[j]	
Middle Eastern	2 (0.8)	2 (1.8)[i]	2 (1.4)	2 (1.4)[j]	
Mixed heritage/other(s)	3 (1.2)	3 (2.7)[i]	8 (5.4)	8 (5.6)[j]	
Prefer not to answer	6 (2.4)	6 (5.4)[i]	6 (4.1)	6 (4.2)[j]	
Missing	141 (55.7)	N/A	3 (2.0)	N/A	
English-speaking ability, n (%)					
Very well	108 (42.7)	108 (96.4)[i]	139 (94.6)	139 (94.6)[k]	.77
Well	3 (1.2)	3 (2.7)[i]	7 (4.8)	7 (4.8)[k]	
Not well	1 (0.4)	1 (0.9)[i]	1 (0.7)	1 (0.7)[k]	
Missing	141 (55.7)	N/A	0 (0)	N/A	
Ontario marginalization index, n (%)					
Marginalized	36 (14.2)	36 (14.9)[l]	16 (10.9)	16 (11.6)[m]	.37
Not marginalized	206 (81.4)	206 (85.1)[l]	122 (83.0)	122 (88.4)[m]	
Missing	11 (4.4)	N/A	9 (6.1)	N/A	

[a]Number of unique patients who responded of 517 video survey responses received. The same patient may be counted multiple times.

[b]Number of unique patients who responded of 209 phone survey responses received. The same patient may be counted multiple times.

[c]*P* value compares the distribution of demographic variables between video and telephone groups.

[d]N/A: not applicable.

[e]*P* value compares <$90,000 with ≥$90,000.

[f]n=108.

[g]n=146.

[h]*P* value compares White with non-White.

[i]n=112.

[j]n=144.

[k]n=147.

[l]n=242.

[m]n=138.

Table 2. Video and phone survey responses.

Question	Video visit responses (n=517)	Phone visit responses (n=209)	*P* value[a]
To what degree did the video or phone visit help you with the health issue for which you needed the appointment?, n (%)			
Not at all helpful	3 (0.6)	1 (0.5)	.14
Not helpful	3 (0.6)	2 (1.0)	
Neutral	17 (3.3)	15 (7.2)	
Somewhat helpful	75 (14.5)	35 (16.8)	
Very helpful	417 (80.7)	154 (73.7)	
Missing	2 (0.4)	2 (1.0)	
What would you have done if you were not able to see your doctor through a video or phone visit?[b], n (%)			
Walk-in clinic	20 (3.9)	2 (1.0)	<.001
Emergency department	32 (6.2)	6 (2.9)	
See/talk to my family doctor	145 (28.1)	58 (27.8)	
Scheduled an in-person visit with this doctor	192 (37.1)	99 (47.4)	
I would not have sought care at that time	198 (38.3)	43 (20.6)	
Missing	0 (0)	1 (0.5)	
How likely are you to recommend video or phone visits to a friend on a scale of 1 to 10? (1 = would not recommend and 10 = would highly recommend), n (%)			
1	1 (0.2)	2 (1.0)	.02
2	1 (0.2)	4 (1.9)	
3	3 (0.6)	0 (0.0)	
4	3 (0.6)	2 (1.0)	
5	20 (3.9)	13 (6.2)	
6	17 (3.3)	11 (5.3)	
7	42 (8.1)	16 (7.7)	
8	68 (13.2)	38 (18.2)	
9	78 (15.1)	26 (12.4)	
10	269 (52.0)	86 (41.2)	
Missing	15 (2.9)	11 (5.3)	
Net promoter score, %	60.2	40.4	N/A[c]
Would you like the option to continue having virtual visits with your health care providers after COVID-19?, n (%)			
No	19 (3.7%)	19 (9.1%)	.004
Not sure	89 (17.2%)	44 (21.1%)	
Yes	405 (78.3%)	146 (69.9%)	
Missing	4 (0.8%)	0 (0%)	

[a]*P* value compares the distribution of survey responses between video and telephone groups.

[b]Multiselect question for video visit survey.

[c]N/A: not applicable.

Table 3. Ordinal logistic regression analysis results.

Variable	Model 1: How likely are you to recommend video or phone visits to a friend on a scale of 1 to 10? (1 = would not recommend and 10 = would highly recommend)[a] (n=255)		Model 2: To what degree did the video or phone visit help you with the health issue for which you needed the appointment?[b] (n=268)	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
Ethnicity (reference: White)				
Non-White	0.52 (0.28-0.99)	.047	0.87 (0.42-1.77)	.69
Prefer not to answer	1.19 (0.35-4.12)	.78	1.33 (0.33-5.39)	.69
Family income (reference: $150,000 CAD or more)				
0 to 29,999	3.28 (0.63-17.09)	.16	1.56 (0.28-8.61)	.61
30,000 to 59,999	0.62 (0.21-1.84)	.39	0.51 (0.15-1.72)	.28
60,000 to 89,999	0.52 (0.19-1.45)	.21	0.74 (0.24-2.33)	.61
90,000 to 119,000	0.67 (0.25-1.76)	.42	0.38 (0.13-1.09)	.07
120,000 to 149,999	1.16 (0.26-5.20)	.84	0.42 (0.10-1.81)	.25
Do not know	1.53 (0.44-5.29)	.50	0.60 (0.17-2.14)	.44
Prefer not to answer	0.52 (0.23-1.15)	.11	0.65 (0.26-1.64)	.36
Gender (reference: Male)				
Female	0.99 (0.48-2.04)	.98	1.70 (0.81-3.58)	.16
Overall marginalization (reference: Not marginalized)				
Marginalized	1.87 (0.78-4.49)	.16	2.15 (0.78-5.89)	.14
Age group (years; reference: 30-39)				
0-18	0.32 (0.05-1.92)	.21	1.29 (0.15-11.28)	.82
19-29	0.63 (0.26-1.54)	.31	0.76 (0.29-1.99)	.57
40-49	0.73 (0.34-1.58)	.42	0.72 (0.31-1.71)	.46
50-59	0.26 (0.11-0.64)	.003	0.53 (0.20-1.40)	.20
60-69	1.19 (0.39-3.63)	.76	1.30 (0.39-4.35)	.67
≥70	0.95 (0.33-2.75)	.92	0.83 (0.28-2.41)	.73

[a]Outcome categories: “1-6,” “7-8,” “9-10.”

[b]Outcome categories: “not helpful or neutral,” “somewhat helpful,” “very helpful.”

[c]OR: odds ratio.

Provider Experience

A total of 75 providers completed the survey in June 2020, and 94 providers completed the survey in June 2021 (Table 4). The top 3 survey respondents among providers in 2020 were physicians (47/75, 63%), social workers (7/75, 9%), and psychotherapists (6/75, 8%), while the top 3 provider survey respondents in 2021 were physicians (48/94, 51%), social workers (9/94, 10%), and physiotherapists (8/94, 9%). In both 2020 and 2021, most providers who delivered virtual visits had been practicing for 10 or more years (40/75, 53% in 2020 and 61/94, 65% in 2021).

Responses to the provider experience surveys are shown in Table 5. When asked whether the quality of the virtual visit was similar to that of an in-person visit, 13% (10/75) selected agree or strongly agree in 2020, compared with 28% (26/94) in 2021. In 2020, 67% (50/75) of providers felt that video visits enabled them to sufficiently address their patient’s clinical need compared with 70% (66/94) in 2021. Most providers planned to continue using video visits after the need for physical distancing decreased (53/75, 71% in 2020 and 69/94, 73% in 2021). When asked to rate on a scale of 1 to 10 their likelihood of recommending other providers to do virtual visits for patients, most providers rated a score of 8 or above in 2020 (47/75, 63% for video and 46/75, 61% for telephone) and in 2021 (62/94, 66% for video and 49/94, 52% for telephone). The net promoter scores for video visits were 17.8% in 2020 and 30.4% in 2021, while the net promoter scores for telephone visits were 19.2% in 2020 and 1.1% in 2021.

Table 4. Baseline characteristics of providers who delivered at least one virtual visit.

Variable	June 2020 (n=75), n (%)	June 2021 (n=94), n (%)
Provider type		
Dietitian	0 (0)	2 (2)
Kinesiologist	2 (3)	2 (2)
Nurse	2 (3)	7 (8)
Nurse practitioner	2 (3)	5 (5)
Occupational therapist	2 (3)	2 (2)
Pharmacist	0 (0)	1 (1)
Physician	47 (63)	48 (51)
Physiotherapist	4 (5)	8 (9)
Psychologist	1 (1)	2 (2)
Psychotherapist	6 (8)	5 (5)
Social service worker	1 (1)	3 (3)
Social worker	7 (9)	9 (10)
Other	1 (1)	0 (0)
Years in practice		
1-2	9 (12)	9 (10)
3-5	7 (9)	10 (11)
6-7	12 (16)	6 (6)
8-9	6 (8)	7 (8)
≥10	40 (53)	61 (65)
Missing	1 (1)	1 (1)

Table 5. Provider experience survey responses.

Question	June 2020 (n=75)	June 2021 (n=94)
The quality of examination virtually was similar to an in-person exam., n (%)		
Strongly disagree	11 (15)	11 (12)
Disagree	30 (40)	21 (22)
Neutral	11 (15)	17 (18)
Agree	9 (12)	20 (21)
Strongly agree	1 (1)	6 (6)
Missing	13 (17)	19 (20)
The video visit enabled me to sufficiently address the patient's clinical need., n (%)		
Strongly disagree	0 (0)	2 (2)
Disagree	7 (9)	3 (3)
Neutral	12 (16)	15 (16)
Agree	41 (55)	38 (40)
Strongly agree	9 (12)	28 (30)
Missing	6 (8)	8 (9)
I spent the same amount of time on the video visit as I would have for an in-person visit., n (%)		
Strongly disagree	4 (5)	3 (3)
Disagree	16 (21)	17 (18)
Neutral	7 (9)	6 (6)
Agree	28 (37)	35 (37)
Strongly agree	13 (17)	24 (26)
Missing	7 (9)	9 (10)
I spent the same amount of effort on the video visit as I would have for an in-person visit., n (%)		
Strongly disagree	4 (5)	6 (6)
Disagree	29 (39)	22 (23)
Neutral	11 (15)	12 (13)
Agree	17 (23)	27 (29)
Strongly agree	8 (11)	16 (17)
Missing	6 (8)	11 (12)
I feel I can deliver the same quality care using video visits as in person., n (%)		
Strongly disagree	1 (1)	9 (10)
Disagree	23 (31)	12 (13)
Neutral	23 (31)	19 (20)
Agree	24 (32)	35 (37)
Strongly agree	2 (3)	18 (19)
Missing	2 (3)	1 (1)
I feel I can deliver the same quality care using phone visits as in person., n (%)		
Strongly disagree	2 (3)	9 (10)
Disagree	26 (35)	22 (23)
Neutral	23 (31)	21 (22)
Agree	22 (29)	27 (29)
Strongly agree	2 (3)	15 (16)
Missing	0 (0)	0 (0)

Question	June 2020 (n=75)	June 2021 (n=94)
I plan to continue using video visits after the need for physical distancing decreases., n (%)		
Strongly disagree	0 (0)	4 (4)
Disagree	6 (8)	4 (4)
Neutral	16 (21)	15 (16)
Agree	31 (41)	24 (26)
Strongly agree	22 (29)	45 (48)
Missing	0 (0)	2 (2)
On a scale of 1 to 10, how likely are you to recommend other providers like yourself do video visits for patients?, n (%)		
1	1 (1)	2 (2)
2	0 (0)	1 (1)
3	1 (1)	1 (1)
4	2 (3)	2 (2)
5	3 (4)	5 (5)
6	4 (5)	6 (6)
7	15 (20)	13 (14)
8	23 (31)	17 (18)
9	10 (13)	15 (16)
10	14 (19)	30 (32)
Missing	2 (3)	2 (2)
Net promoter score, %	17.8	30.4
On a scale of 1 to 10, how likely are you to recommend other providers like yourself do phone visits for patients?, n (%)		
1	0 (0)	3 (3)
2	1 (1)	3 (3)
3	0 (0)	3 (3)
4	0 (0)	3 (3)
5	4 (5)	12 (13)
6	7 (9)	10 (11)
7	15 (20)	11 (12)
8	20 (27)	14 (15)
9	11 (15)	9 (10)
10	15 (20)	26 (28)
Missing	2 (3)	0 (0%)
Net promoter score, %	19.2	1.1

Discussion

Principal Findings

This study describes the patient and provider experiences with virtual visits during the COVID-19 pandemic across an academic ambulatory hospital in Toronto, Canada. Feedback for virtual visits was generally positive among patients. Video visits were the preferred modality over telephone among many patients, particularly for those who were new patients. However, we found that patients of non-White background were less likely to recommend virtual visits compared with those of White background. Provider experiences with virtual visits were less positive compared with those of their patients, but there was a general improvement in provider feedback from 2020 to 2021.

Comparison With Prior Work

Most patients found that their virtual visit was helpful in addressing their health issue and rated a high score when asked to what degree they would recommend virtual visits to a friend. However, a higher proportion of patients reported video visits to be “very helpful” compared with telephone visits. Similarly, the net promoter score was much higher for video visits compared with telephone visits. In our study, patients who

completed telephone visit surveys were generally female and older in age compared with patients who completed video visits. In the literature, older patients are less likely to engage in virtual care than in in-person care and even less likely to choose video than telephone [24,25].

In our study, several patients indicated in the open-ended questions that they would prefer video over telephone due to the ability to see the provider and observe facial expressions and body language. Other studies have also cited the benefit of increased human connection that accompanies video platforms [26]. The difference in net promoter scores was even greater among new patients, with video visits reaching a significantly higher score compared with telephone visits. Video visits enable patients to see their provider, which supports the development of a patient-provider relationship especially for an initial encounter. These results contrast with findings from a systematic review that reported no significant differences in patient satisfaction between video and telephone visits, but they did not stratify initial versus follow-up encounters [27]. This may also be attributed to our sample consisting of younger individuals. It is likely that younger patients prefer video visits more so than older patients who may prefer telephone visits due to ease of access [28].

Despite the generally positive feedback for virtual visits, a small proportion of patients did not find the visit helpful or rated a low recommendation score. The open-ended responses suggest that some patients were unhappy with the delay in their appointment start time and the lack of communication from the clinic in such cases. Published studies have cited other patient criticisms of virtual visits such as technical issues with connection and quality of the call [29], a lack of privacy at home when attending virtual visits [30], and a preference for in-person visits for certain physical health issues or to build a relationship with their provider [31]. Overall, it appears that many patients had positive experiences with their virtual visit(s), with several citing reasons such as convenience and that they were able to save time and money [13].

Findings from the regression model indicate that non-White patients were less likely to recommend virtual visits to a friend compared with White patients. Mixed findings are reported in the literature, with several studies showing no significant differences in patient experience with virtual care by ethnicity [32-34], while others have shown that patients of non-White backgrounds are more likely to have lower satisfaction with virtual visits compared with their White counterparts [17,35]. Upon analysis of the average recommendation scores for video versus telephone visits by ethnicity, shown in Table S5 in Multimedia Appendix 2, Asian and Black patients had similar, if not better, scores than White patients; however, ratings were generally lower for the other ethnic groups. Reasons for this disparity remain unclear and should be investigated in future work. Among all age groups of interest, only patients aged 50 years to 59 years were found to be less likely to recommend virtual visits than patients aged 30 years to 39 years (the reference group). This may be explained by older patients' preferences toward in-person care or the technological barriers they may encounter with virtual visits [36,37]; however, this association did not persist in the older age groups (those older than 60 years) for reasons unknown. We do note that there were no significant differences in experience found for the other demographic variables assessed (family income, overall marginalization, sex, and most age groups), which may be a positive sign that the delivery of virtual visits may have helped to bridge the gap in equitable health care access in certain ways; for example, lower income patients may find it easier to attend a virtual visit than request time off work to attend in-person care, or older patients with mobility issues may find it easier to attend a virtual visit than an in-person visit.

Our findings are consistent with other studies that found lower satisfaction with virtual care among health care providers compared with patients [38]. Our findings show that the net promoter scores were lower for both video and telephone visits when rated by providers than by patients. Open-ended responses from the provider surveys suggest that many providers felt that they needed to provide a physical examination to adequately address their patient's health needs, similar to findings from the available literature [39]. Other providers felt that the quality of the virtual visit was lower than that of an in-person visit. Several also cited technical issues, particularly with video visits, as a deterrent for virtual care. Studies examining provider experience with virtual care reported similar reasons for provider dissatisfaction [40], with less than one-half of providers preferring virtual over in-person care [41].

From a provider standpoint, there appeared to be an increase in positive feedback for virtual visits across most survey questions from 2020 to 2021, including quality of virtual visit, time and effort spent on virtual visit, and preference to use virtual care after the pandemic. This may be because, as providers had more experience with virtual care, their self-efficacy may have improved. Another explanation is that the proportion of all visits that were virtual was higher in 2020 than in 2021, so the appropriateness of virtual care for the visit reason was likely also higher in 2021. Several studies have reported that clinicians have a positive outlook on virtual visits [13], particularly within the mental health field [39,42] in which physical examinations play a lesser role in clinical practice compared with other specialties. However, we note that the providers' net promoter score for telephone visits decreased significantly from 2020 to 2021, while scores for video visits increased. A possible explanation supported by open-text portions of the survey is that providers may have been more comfortable and proficient with providing video visits and preferred the face-to-face connection that can be missing from telephone communication, but this merits further exploration.

Strengths and Limitations

A strength of this study is that it provides the patient and provider perspectives on virtual care in a large ambulatory hospital setting with responses across many clinical specialties and programs. Our findings also offer insights into both patient and provider experiences and into some of the demographic differences in experience with virtual care to identify potential equity issues. Nonetheless, our study does have several limitations. The overall response rate for demographic questions in the video survey was lower than anticipated due to a technical error in survey deployment. Furthermore, despite our equity

focus, there are limitations in the demographic insights that can be gleaned. First, although we captured several important demographic variables in this study, we were unable to assess the association between patient experience and other potentially relevant characteristics, such as education level, employment, and immigration status. The low proportion of patients who consented to receive surveys and the fact that surveys could only be sent to patients who had a valid email address and were registered on the portal system would have limited responses, including from certain marginalized and underserviced groups. We also acknowledge the possibility that patients who have fewer positive experiences with virtual visits may be less inclined to complete the survey, which would potentially bias the findings to be more positive. However, our analysis was still able to detect a difference in experience among patients of ethnic minorities. Finally, an electronic survey does not offer a deep understanding of experience, particularly among marginalized groups, as opportunities for feedback is limited and patients may not feel as comfortable sharing their thoughts on the platform due to confidentiality concerns. Future studies of patient experiences with virtual care should include interviews or focus groups with patients from underserved communities. We also surveyed patients seen in specialty clinics within an ambulatory care hospital and did not include primary care patient surveys. Last, these findings reflect the experiences of patients and providers in a single institution within a universal health care system and therefore may not be generalizable to other settings.

Conclusions

This study summarizes the patient and provider experiences with virtual care across an academic ambulatory hospital in Toronto, Canada during the COVID-19 pandemic. Virtual care, comprised of video and telephone visits, was generally well-received among most patients, with many favoring video visits over telephone visits especially for new patients. However, virtual care was less endorsed among many providers. Furthermore, patients with non-White ethnic backgrounds were less likely to recommend virtual visits. These findings provide important contributions regarding understanding overall patient and provider experiences with virtual care as well as predictors of patient experience. Given the prospect of the hybrid modality of care delivery that includes both virtual and in-person options of care delivery post-COVID-19, future work should aim to develop ways to understand factors to improve patient and provider experiences with virtual care and to assess the impact of virtual modalities on patient outcomes and quality of care.

Acknowledgments

The authors want to acknowledge the following individuals who provided support for this work: Women’s Virtual team for providing feedback on the survey design; Patricia Rios and Hayley Baranek for Women’s College Hospital Institute for Health System Solutions and Virtual Care (WIHV) project management support; Suman Budhwani and Ian McMillan for supporting the development of the surveys; the decision support team and the information management (IM)/information technology (IT) team for supporting development of the consent workflow and survey deployment; and Drew Wesley for providing support and feedback for this study.

There was no direct funding provided for this study. PA is funded in part by a New Investigator Award from the Department and Community Medicine at the University of Toronto.

Authors' Contributions

CC led and conducted the analysis and initial drafting of the manuscript. GM led the conception of the study, supervised the overall design and activities of the evaluation, and supervised the drafting and editing of the manuscript. DN supported the development of the surveys and program evaluation. PA co-led the conceptualization of the overall virtual program evaluation. OB and DM provided feedback on overall program evaluation and supported manuscript writing. All authors read, edited, and approved the final manuscript prior to publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and provider survey questions.
[DOCX File , 29 KB - jmir_v24i10e38604_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 27 KB - jmir_v24i10e38604_app2.docx]

References

1. Using Telehealth to Expand Access to Essential Health Services during the COVID-19 Pandemic. Centers for Disease Control and Prevention. 2020 Jun 10. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2022-10-08]
2. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358 [FREE Full text] [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
3. Mehrotra A, Bhatia RS, Snoswell CL. Paying for telemedicine after the pandemic. JAMA 2021 Feb 02;325(5):431-432 [FREE Full text] [doi: 10.1001/jama.2020.25706] [Medline: 33528545]
4. Re: Virtual Care Program - Billing Amendments to Enable Direct-to-Patient Video Visits and Modernize Virtual Care Compensation. Ontario Ministry of Health and Long-Term Care. 2019 Nov 15. URL: https://www.health.gov.on.ca/en/pro/programs/ohip/bulletins/4000/bul4731.aspx [accessed 2022-10-08]
5. Re: Changes to the Schedule of Benefits for Physician Services (Schedule) in response to COVID-19 influenza pandemic effective March 14, 2020. Ontario Ministry of Health and Long-Term Care. 2020 Mar 13. URL: https://www.health.gov.on.ca/en/pro/programs/ohip/bulletins/4000/bul4745.aspx [accessed 2022-10-08]
6. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: a repeated cross-sectional study. CMAJ Open 2021 Feb 17;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
7. Itamura K, Rimell FL, Illing EA, Higgins TS, Ting JY, Lee MK, et al. Assessment of patient experiences in otolaryngology virtual visits during the COVID-19 pandemic. OTO Open 2020;4(2):2473974X20933573 [FREE Full text] [doi: 10.1177/2473974X20933573] [Medline: 32551407]
8. Gold KJ, Laurie AR, Kinney DR, Harmes KM, Serlin DC. Video visits: family physician experiences with uptake during the COVID-19 pandemic. Fam Med 2021 Mar 4;53(3):207-210. [doi: 10.22454/fammed.2021.613099]
9. Lavery MJ, Phillips D, Yip V, Azurdia R, Thompson B. Survey to evaluate the patient experience of virtual telephone consultations during the COVID-19 pandemic. Clin Dermatol 2021;39(2):334-336 [FREE Full text] [doi: 10.1016/j.clindermatol.2020.11.004] [Medline: 34272032]
10. Banks J, Corrigan D, Grogan R, El-Naggar H, White M, Doran E, et al. LoVE in a time of CoVID: Clinician and patient experience using telemedicine for chronic epilepsy management. Epilepsy Behav 2021 Feb;115:107675 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107675] [Medline: 33342712]
11. Björndell C, Premberg Å. Physicians' experiences of video consultation with patients at a public virtual primary care clinic: a qualitative interview study. Scand J Prim Health Care 2021 Mar;39(1):67-76 [FREE Full text] [doi: 10.1080/02813432.2021.1882082] [Medline: 33650941]
12. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 May;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
13. Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]
14. Appireddy R, Khan S, Leaver C, Martin C, Jin A, Durafourt BA, et al. Home virtual visits for outpatient follow-up stroke care: cross-sectional study. J Med Internet Res 2019 Oct 07;21(10):e13734 [FREE Full text] [doi: 10.2196/13734] [Medline: 31593536]
15. Chan S, O'Riordan A, Appireddy R. Exploring the determinants and experiences of senior stroke patients with virtual care. Can J Neurol Sci 2021 Jan 27;48(1):87-93. [doi: 10.1017/cjn.2020.162] [Medline: 32713397]
16. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
17. Serper M, Nunes F, Ahmad N, Roberts D, Metz DC, Mehta SJ. Positive early patient and clinician experience with telemedicine in an academic gastroenterology practice during the COVID-19 pandemic. Gastroenterology 2020 Oct;159(4):1589-1591.e4 [FREE Full text] [doi: 10.1053/j.gastro.2020.06.034] [Medline: 32565015]
18. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
19. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
20. 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]
21. Ontario Marginalization Index (ON-Marg). Public Health Ontario. 2016. URL: https://www.publichealthontario.ca/en/data-and-analysis/health-equity/ontario-marginalization-index [accessed 2022-10-08]
22. Reichheld F. The One Number You Need to Grow. Harvard Business Review. 2003 Dec. URL: https://hbr.org/2003/12/the-one-number-you-need-to-grow [accessed 2022-10-08]
23. What is a good Net Promoter Score? Qualtrics XM. URL: https://www.qualtrics.com/experience-management/customer/good-net-promoter-score/ [accessed 2022-10-08]

24. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient characteristics associated with choosing a telemedicine visit vs office visit with the same primary care clinicians. JAMA Netw Open 2020 Jun 01;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]
25. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
26. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. Lancet Digit Health 2020 Jun;2(6):e282-e285 [FREE Full text] [doi: 10.1016/S2589-7500(20)30104-7] [Medline: 32382724]
27. Rush KL, Howlett L, Munro A, Burton L. Videoconference compared to telephone in healthcare delivery: A systematic review. Int J Med Inform 2018 Oct;118:44-53. [doi: 10.1016/j.ijmedinf.2018.07.007] [Medline: 30153920]
28. Schifeling CH, Shanbhag P, Johnson A, Atwater RC, Koljack C, Parnes BL, et al. Disparities in video and telephone visits among older adults during the COVID-19 pandemic: cross-sectional analysis. JMIR Aging 2020 Nov 10;3(2):e23176 [FREE Full text] [doi: 10.2196/23176] [Medline: 33048821]
29. Wade VA, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. BMC Health Serv Res 2010 Aug 10;10(1):233 [FREE Full text] [doi: 10.1186/1472-6963-10-233] [Medline: 20696073]
30. Fletcher S, Tsang V. The era of virtual care: Perspectives of youth on virtual appointments in COVID-19 and beyond. Paediatr Child Health 2021 Jul;26(4):210-213 [FREE Full text] [doi: 10.1093/pch/pxaa138] [Medline: 34127935]
31. Krishnamurthy Y, Pagliaro J, Grady C, Katz N, Bunn D, Bhatt A. Patient evaluation of a virtual visit program for adults with congenital heart disease. Am Heart J 2021 Dec;242:138-145 [FREE Full text] [doi: 10.1016/j.ahj.2021.08.004] [Medline: 34400141]
32. Johnson BA, Lindgren BR, Blaes AH, Parsons HM, LaRocca CJ, Farah R, et al. The new normal? Patient satisfaction and usability of telemedicine in breast cancer care. Ann Surg Oncol 2021 Oct 17;28(10):5668-5676 [FREE Full text] [doi: 10.1245/s10434-021-10448-6] [Medline: 34275045]
33. Shin C, Allen A, Zhu D, Tellechea L, Watts K, Abraham N. Patient satisfaction and savings, and clinical outcomes of televisits in female pelvic medicine and reconstructive surgery at an urban academic center. Neurourol Urodyn 2021 Sep;40(7):1834-1844 [FREE Full text] [doi: 10.1002/nau.24759] [Medline: 34342368]
34. Gustke SS, Balch DC, West VL, Rogers LO. Patient satisfaction with telemedicine. Telemedicine Journal 2000 May;6(1):5-13. [doi: 10.1089/107830200311806]
35. Myburgh N, Solanki G, Smith M, Lalloo R. Patient satisfaction with health care providers in South Africa: the influences of race and socioeconomic status. Int J Qual Health Care 2005 Dec;17(6):473-477. [doi: 10.1093/intqhc/mzi062] [Medline: 15985504]
36. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
37. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. JMIR Med Inform 2020 Aug 12;8(8):e20359 [FREE Full text] [doi: 10.2196/20359] [Medline: 32784177]
38. Barkai G, Gadot M, Amir H, Menashe M, Shvimer-Rothschild L, Zimlichman E. Patient and clinician experience with a rapidly implemented large-scale video consultation program during COVID-19. Int J Qual Health Care 2021 Feb 20;33(1):1 [FREE Full text] [doi: 10.1093/intqhc/mzaa165] [Medline: 33313891]
39. Connolly SL, Gifford AL, Miller CJ, Bauer MS, Lehmann LS, Charness ME. Provider perceptions of virtual care during the coronavirus disease 2019 pandemic: a multispecialty survey study. Med Care 2021 Jul 01;59(7):646-652 [FREE Full text] [doi: 10.1097/MLR.0000000000001562] [Medline: 34009880]
40. Nies S, Patel S, Shafer M, Longman L, Sharif I, Pina P. Understanding physicians' preferences for telemedicine during the COVID-19 pandemic: cross-sectional study. JMIR Form Res 2021 Aug 13;5(8):e26565 [FREE Full text] [doi: 10.2196/26565] [Medline: 34227993]
41. Malouff T, TerKonda S, Knight D, Abu Dabrh AM, Perlman A, Munipalli B, et al. Physician satisfaction with telemedicine during the COVID-19 pandemic: the Mayo Clinic Florida experience. Mayo Clin Proc Innov Qual Outcomes 2021 Aug;5(4):771-782 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2021.06.006] [Medline: 34226884]
42. Gentry MT, Puspitasari AJ, McKean AJ, Williams MD, Breitinger S, Geske JR, et al. Clinician satisfaction with rapid adoption and implementation of telehealth services during the COVID-19 pandemic. Telemed J E Health 2021 Dec 19;27(12):1385-1392. [doi: 10.1089/tmj.2020.0575] [Medline: 33606560]

Abbreviations

IM: information management
IT: information technology
ON-Marg: Ontario marginalization index

OR: odds ratio
WCH: Women's College Hospital
WIHV: Women's College Hospital Institute for Health System Solutions and Virtual Care

Edited by C Basch; submitted 11.04.22; peer-reviewed by M Lee, E Serhal, D Madhusudhan; comments to author 12.07.22; revised version received 25.07.22; accepted 22.08.22; published 25.10.22.

Please cite as:
Chu C, Nayyar D, Bhattacharyya O, Martin D, Agarwal P, Mukerji G
Patient and Provider Experiences With Virtual Care in a Large, Ambulatory Care Hospital in Ontario, Canada During the COVID-19 Pandemic: Observational Study
J Med Internet Res 2022;24(10):e38604
URL: https://www.jmir.org/2022/10/e38604
doi:10.2196/38604
PMID:36194862

Original Paper

Sexually Transmitted Disease–Related Reddit Posts During the COVID-19 Pandemic: Latent Dirichlet Allocation Analysis

Amy K Johnson[1,2*], MSW, PhD; Runa Bhaumik[3*], PhD; Debarghya Nandi[3], MS; Abhishikta Roy[3], MS; Supriya D Mehta[4*], PhD

[1]The Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]School of Public Health, University of Illinois at Chicago, Chicago, IL, United States

[4]Rush University, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Amy K Johnson, MSW, PhD
The Potocsnak Family Division of Adolescent and Young Adult Medicine
Ann & Robert H. Lurie Children's Hospital of Chicago
225 E Chicago
Box 161
Chicago, IL, 60611
United States
Phone: 1 3122277733
Email: akjohnson@luriechildrens.org

Abstract

Background: Sexually transmitted diseases (STDs) are common and costly, impacting approximately 1 in 5 people annually. Reddit, the sixth most used internet site in the world, is a user-generated social media discussion platform that may be useful in monitoring discussion about STD symptoms and exposure.

Objective: This study sought to define and identify patterns and insights into STD-related discussions on Reddit over the course of the COVID-19 pandemic.

Methods: We extracted posts from Reddit from March 2019 through July 2021. We used a topic modeling method, Latent Dirichlet Allocation, to identify the most common topics discussed in the Reddit posts. We then used word clouds, qualitative topic labeling, and spline regression to characterize the content and distribution of the topics observed.

Results: Our extraction resulted in 24,311 total posts. Latent Dirichlet Allocation topic modeling showed that with 8 topics for each time period, we achieved high coherence values (pre–COVID-19=0.41, prevaccination=0.42, and postvaccination=0.44). Although most topic categories remained the same over time, the relative proportion of topics changed and new topics emerged. Spline regression revealed that some key terms had variability in the percentage of posts that coincided with pre–COVID-19 and post–COVID-19 periods, whereas others were uniform across the study periods.

Conclusions: Our study's use of Reddit is a novel way to gain insights into STD symptoms experienced, potential exposures, testing decisions, common questions, and behavior patterns (eg, during lockdown periods). For example, reduction in STD screening may result in observed negative health outcomes due to missed cases, which also impacts onward transmission. As Reddit use is anonymous, users may discuss sensitive topics with greater detail and more freely than in clinical encounters. Data from anonymous Reddit posts may be leveraged to enhance the understanding of the distribution of disease and need for targeted outreach or screening programs. This study provides evidence in favor of establishing Reddit as having feasibility and utility to enhance the understanding of sexual behaviors, STD experiences, and needed health engagement with the public.

(J Med Internet Res 2022;24(10):e37258) doi:10.2196/37258

KEYWORDS

infodemiology; Latent Dirichlet Allocation; natural language processing; Reddit; sexually transmitted infections; surveillance; social media; COVID-19; social media content; content analysis; health outcome; infoveillance; health information; sexually transmitted disease; STD

Introduction

More than 2.5 million cases of chlamydia, gonorrhea, and syphilis were reported in 2019, with sexually transmitted diseases (STD) cases reaching an all-time high for the sixth consecutive year in the United States [1]. STDs are common and costly, impacting approximately 1 in 5 people annually and accounting for US $16 billion in annual health care costs [2]. New data from the Centers for Disease Control and Prevention demonstrate that during the start of the COVID-19 pandemic (from March to April 2020), reported STD cases dramatically decreased compared to the same time in 2019. At that point, the current cumulative totals for STD cases compared to 2019 were 1% lower for primary and secondary syphilis, 7% lower for gonorrhea, and 14% lower for chlamydia [3]. Although case reports were lower for the first part of 2020, cases rebounded later in the year and were on track to surpass 2019 totals [3].

Multiple factors likely contributed to the observed decrease in reported STD cases during the early phases of the COVID-19 pandemic. Restrictions of in-person clinic visits resulted in reduced screening of asymptomatic patients. The Centers for Disease Control and Prevention provided guidance for sexual health services to prioritize patients based on symptoms and risk, along with delaying routine screening until after the emergency response [4]. Many health department staff were redeployed from STD tracking to COVID-19 contact tracing and control [5]; 57% of disease intervention specialists reported that they were reassigned from STD to COVID-19 services, limiting the workforce available to provide STD prevention, screening, and treatment [5]. Finally, national stay-at-home orders were issued during phases of the pandemic that were designed to reduce the spread of COVID-19 but may also have reduced STD transmission by reducing sexual behavior outside of the household, limiting the number of new sexual partners, and restricting sexual networks [6].

Recent estimates indicate that 80% of all internet users report accessing health information on the web [7]. As the internet can be accessed anonymously and at any time, users can seek STD information and resources confidentially, which may facilitate more frequent and open disclosure of symptoms and exposure experiences [8]. Reddit, the sixth most used internet site in the world, is a user-generated social media discussion platform that may be useful in monitoring discussion about STD symptoms and exposure [9]. Reddit is considered one of the most authentic web spaces as there are safeguards against “bot accounts” and rich communication occurs without the barrier of requiring demographic or identifying information to join [10,11]. Prior health research has established that Reddit is an acceptable platform to conduct scientific investigations [10,12,13]. Topic-specific Reddit discussions (subreddits) dedicated to discussing sexual health and STDs may provide valuable insight to exposure, symptoms, testing, and sexual behavior during the COVID-19 pandemic. Prior analyses of Reddit discussion content have been conducted across different diseases and health conditions, including smoking cessation, atopic dermatitis, suicide, and pregnancy [10-13]. To derive meaningful and replicable information from Reddit discussion content, the complexity of high-volume text data needs numerical structure implemented in an unbiased way. Latent Dirichlet Allocation (LDA) is a natural language processing method that identifies common words and topics in text and allows experts to assess common themes among findings [14]. This study sought to define and identify patterns and insights into STD-related discussions on Reddit via LDA over the course of the COVID-19 pandemic. Our team hypothesized that there would be an increase in the volume of STD-related posts on Reddit and the variation of topics during the COVID-19 pandemic compared to the prepandemic period due to behavior changes during the COVID-19 pandemic.

Methods

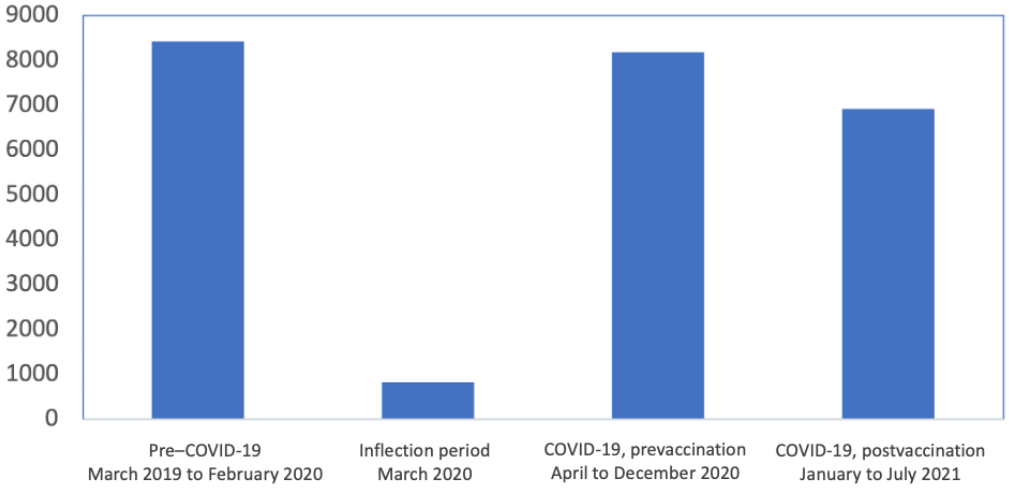
Ethics Approval

The study protocol was determined to be nonhuman subjects research by the Ann & Robert H. Lurie Children’s Hospital Institutional Review Board (IRB#2022-4964) because of the use of publicly available, nonidentifiable data.

Data Extraction

This study used publicly available data from the web-based discussion forum, Reddit. Reddit is an anonymous social media site that is user-generated and discussion-based. The site is organized into “subreddits” that are content-specific. Posts were extracted from 2 subreddits: “STD” (r/STD) and “sexual health”(r/sexualhealth). However, due to the small number of posts in r/sexual health, we only used the subreddit r/STD in our analysis. The *pushshift* Reddit application programming interface was used for searching Reddit comments and submissions [15]. Reddit’s official application programming interface (Reddit 2021) was used to collect posts and associated metadata (date) from r/STD and r/sexualhealth from March 2019 to July 2021, resulting in 24,311 posts [10]. Only English-language posts were included in the analysis. Figure 1 displays the number of posts that were extracted from each subreddit for the time frames used in the analysis. “Pre–COVID-19” was defined as ranging from March 2019 to February 2020 (8421 posts); “COVID-19, prevaccination” was defined as ranging from April to December 2020 (8169 posts); “COVID-19, postvaccination” was defined as ranging from January to July 2021 (6908 posts); and an inflection period was defined as March 2020 (813 posts). Based on the most current literature, we reasonably supposed the absence of seasonality in sexually transmitted infection (STI) cases [16,17].

Figure 1. Distribution of volume of posts for study time periods from March 2019 to July 2021, resulting in a total of 24,311 posts.



Data Preprocessing

Data preprocessing steps were conducted following common approaches in natural language processing [18]. Preprocessing eliminates some of the inconsistencies in the data and reduces the content to useable text. In all, 4 preprocessing steps were completed on each line from the text file to extract and clean each title, body, and comment separately: (1) the removal of URLs, (2) tokenization, (3) punctuation and stop word removal, and (4) lemmatization [19-21].

Statistical Analysis

LDA Topic Modeling

We used an increasingly popular topic modeling method, LDA, to conduct a text analysis identifying the most common topics discussed in the Reddit posts [22]. LDA is a statistical generative model that discovers latent semantic topics in large collections of text documents (posts in our study), where each document results from random mixtures over latent topics and each topic is characterized by a distribution over words. The model is presented in plate notation in Figure 2 [14]. Both the topics and words have a Dirichlet prior distribution, respectively, with α being the parameter of the per-document Dirichlet prior on the topics, and β being the parameter of the per-word Dirichlet prior on the words. θ_m is the topic distribution for document m. φ_k is the word distribution for topic k. Z_{nm} is the topic for the nth word in the mth document. W_{nm} is the actual nth word in the mth document. Considering the nature of its structure, LDA is a multiple-level hierarchical Bayesian model.

To conduct the LDA, we converted the corpus to a document-term matrix, comprising rows representing original posts and columns representing each word in the corpus. Each cell in the document-term matrix contains the frequency of times a specific word (defined by the column) occurred in a specific post (defined by the row). From this document-term matrix, the entire corpus was represented, including patterns of words that commonly occur together within the same post. We used the *gensim* library to perform LDA model estimation, which determined sets of words that appeared frequently together in posts across sexual health subreddits [19].

The LDA model then outputs a topic-document matrix, representing the relative importance of each topic in each document. Models were applied to pre–COVID-19 posts from March 2019 to February 2020 (8421 posts), prevaccination posts from April to December 2020 (8169 posts), and postvaccination posts from January to July 2021 (6908 posts; Figure 3). For topic modeling, we excluded posts for the inflection period (March 2020; 813 posts).

A key process in LDA is to estimate the optimal number of topics. To estimate the number of topics, we used the topic coherence index, which is the most consistent measure of human interpretability [23]. Topic coherence measures score a single topic by measuring the degree of semantic similarity between high-scoring words in the topic. These measurements help distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference. The higher the topic coherence score, the better the quality of the model. To avoid overfit and sparsity and improve inference, we selected the number of topics as 8. Topics were reviewed and labeled independently by 2 experts in STD epidemiology and control (AKJ and SDM). Once independent review was completed, labels were discussed until consensus was reached, resulting in 100% agreement.

Figure 2. Latent Dirichlet Allocation in plate notation (adapted from Blei et al [14]).

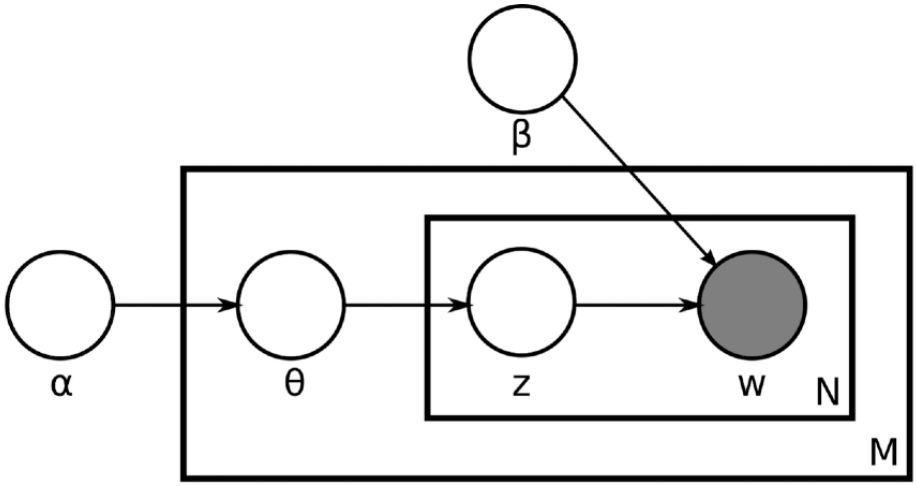
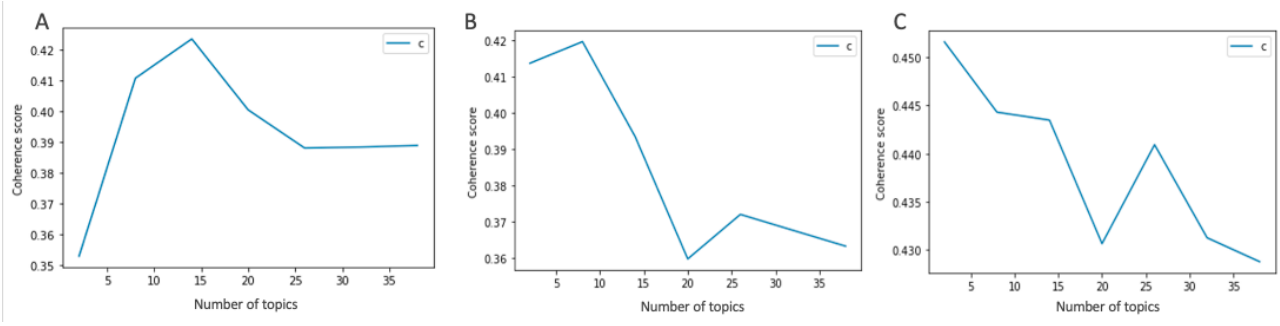


Figure 3. Latent Dirichlet Allocation topic modeling on multiple time periods: (A) pre–COVID-19, (B) prevaccination, and (C) postvaccination.



Word Cloud

A word cloud is a text visualization technique that focuses on the frequency of words and correlates the size and opacity of a word to its frequency within a text body. The output is usually an image that depicts different words in different sizes and opacities relative to the word frequency. Separate frames were created for posts containing the following terms: chlamydia, gonorrhea, syphilis, gonorrhea/discharge/dysuria, and syphilis/chancre/ulcer. After data preprocessing was completed, each string was passed to the *WordCloud* function in Python to generate a word cloud [24]. For *WordCloud* visualization, we chose 3 etiologic terms (chlamydia, gonorrhea, and syphilis) and 3 of the most common terminologies from the topic search: herpes/herpes simplex virus (HSV)/human papillomavirus (HPV; as a single topic, due to correlations), diagnosis/testing, and STI/STD. Each separate word cloud was formed by searching each word in the topic.

Spline Regression Plots

Spline regression modeling has become popular in applied clinical research. Modern biostatistics makes use of spline regression to model smooth functions such as time series, cumulative effects, and frequency distributions and in survival analysis. Spline regression is used to overcome the difficulties of linear and polynomial regression algorithms. In linear regression, the entire data set is considered once. Polynomial regression can express a particular amount of curvature in a nonlinear relationship, but in spline regression (a nonparametric regression), the data set is divided into bins. Each bin of the data is fitted with separate models. The points where the data are divided into bins are called knots. In simpler words, splines are piecewise polynomial functions. To identify patterns in the change of the proportion of posts related to certain search terms relative to the total number of posts in a particular month over the entire study period (spanning from March 2019 to July 2021), a spline plot was created. The pre-COVID-19; inflection; COVID-19, prevaccination; and COVID-19, postvaccination periods were highlighted on the plots for a better understanding of search trends across time. The plots were created using *ggplot2* package in R statistical software (R Foundation for Statistical Computing) [25]. For spline regression, we used a cubic B-spline basis with 2 boundary knots and 1 interior knot placed at the median of the observed data values. As with the word clouds, we created 3 plots based on etiology (chlamydia, gonorrhea, and syphilis) and 3 plots based on common topics (diagnose/test/tested, herpes/HSV/HPV, and gonorrhea/dysuria/discharge). A detailed review of spline regression using R software can be found in Perperoglou et al [26].

Results

Reddit Posts

Of the 24,311 posts, the average number of posts per month were 701.75 during the pre–COVID-19 period; 907.67 during the COVID-19, prevaccination period; and 863.50 during the COVID-19, postvaccination period, but there was substantial variability from month to month and within each time period. The average number of posts per month per period demonstrated

growth in subreddit volume during COVID-19. Figure 4 displays the number of posts per month by observation period. May 2019 consisted of 210 posts and August 2021 consisted of 169 posts, which were 2 of the lowest volumes recorded and were both preceded by 2 months of high-volume posts.

LDA topic modeling showed that with 8 topics for each time period, we achieved high coherence values (pre–COVID-19=0.41, prevaccination=0.42, and postvaccination=0.44). Figure 5 shows the distribution of topic posts in pre–COVID-19, prevaccination, and postvaccination "STD" and "sexual health" subreddits over the 8 topics extracted using LDA. Although most topic categories remained the same over time, the relative proportion of topics changed and new topics emerged. In the pre–COVID-19 period, a general category of "STD Risk" emerged with no specific etiology or mention of symptoms with words such as "negative" and "exposure" in the top 10 terms associated with the topic (Table 1). "HPV" and "warts" as terms did not appear in the pre–COVID-19 period. There was specific language surrounding herpes symptoms (eg, "outbreak") and diagnosis (eg, testing and positive or negative) and the introduction of "HSV" in the postvaccination period, whereas words used in conjunction with herpes in previous periods were primarily related to images and nonspecific symptoms (eg, "redness" and "bumps"; Table 2). Moreover, although the "herpes image" topic category included nonspecific symptoms (eg, "bump" and "redness"), this categorization diverged during COVID-19 periods, with a topic category emerging for penile "bump" without the mention of herpes. In the postvaccination period, the "oral sex/STD questioning" topic included the term "penis"; although this topic existed in the other 2 periods, it did not include "penis" as a top 10 term (Table 3).

Figure 4. Average number of Reddit posts per month, by period.

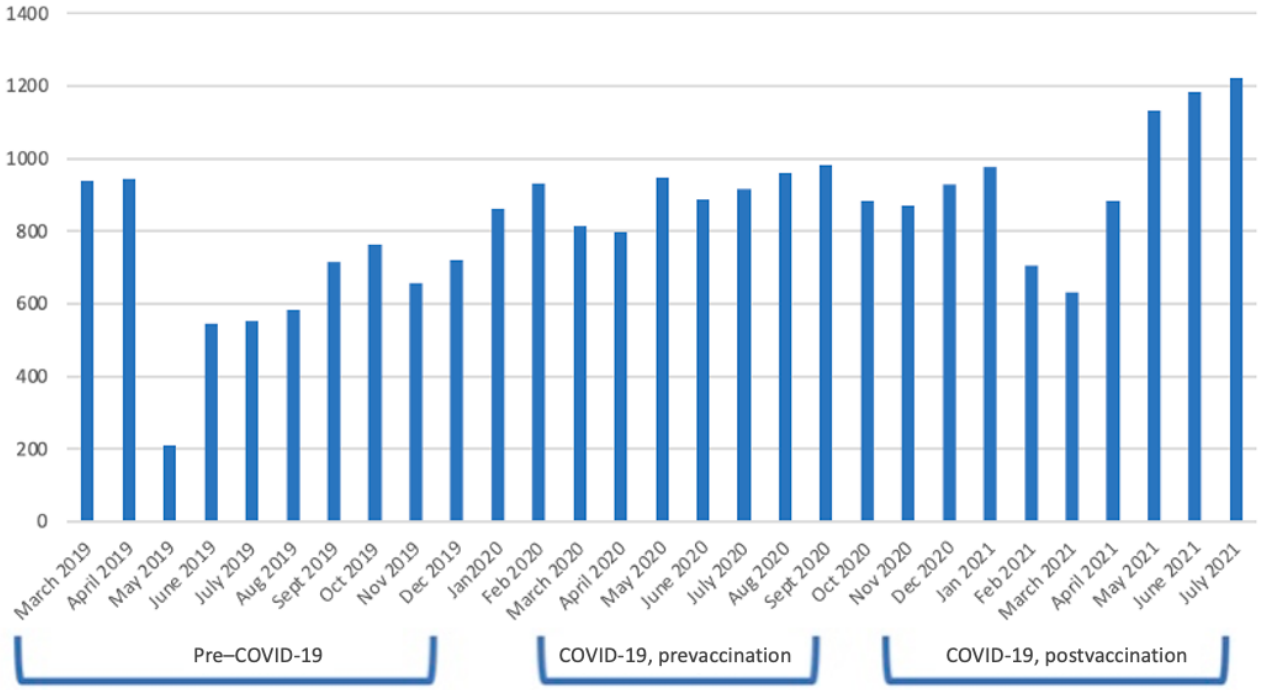


Figure 5. Distribution of posts: the proportion of documents that are assigned to each topic. STD: sexually transmitted disease; HPV: human papillomavirus.

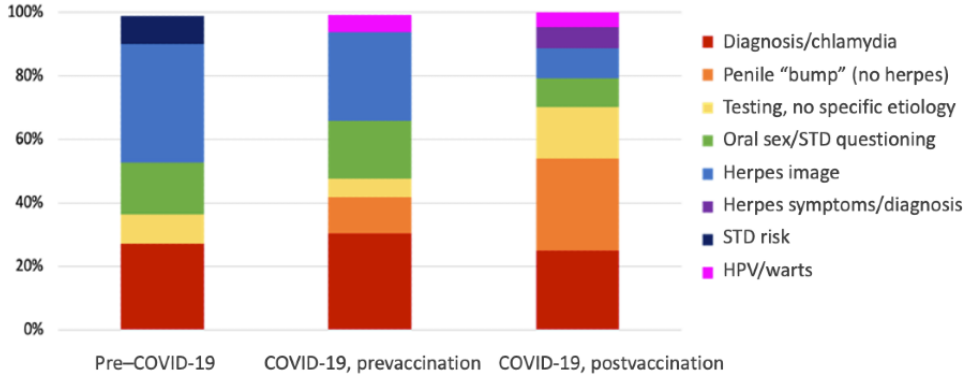


Table 1. Pre–COVID-19 topics and the top 10 terms derived from a Latent Dirichlet Allocation model created from 3 different time periods: pre–COVID-19, prevaccination, and postvaccination.

Topic	Top 10 terms
Testing, no specific etiology	day, test, week, negative, feel, take, pain, still, exposure, also
Herpes image	com, imgur, penis, sex, STD[a], help, really, day, condom, herpes
Oral sex/STD questioning	sex, oral, day, ago, STD, know, condom, unprotected, time, penis
Penile “bump”	com, imgur, bump, look, red, penis, herpes, week, day, spot
Diagnosis/doctor (results)	test, say, back, doctor, look, hepatitis, come, herpes, throat, negative
Diagnosis/chlamydia	test, sex, chlamydia, week, month, symptom, come, back, time, partner

[a]STD: sexually transmitted disease.

Table 2. Prevaccination topics and the top 10 terms derived from a Latent Dirichlet Allocation model created from 3 different time periods: pre–COVID-19, prevaccination, and postvaccination.

Topic	Top 10 terms
Oral sex/STD[a] questioning	sex, oral, come, week, test, day, know, back, STD, time
HPV[b]/warts treatment, herpes questioning	test, sex, HPV, time, year, condom, month, ago, last, say
Diagnosis/chlamydia	test, take, symptom, day, week, know, say, back, chlamydia, doctor
Testing, no specific etiology image	day, month, feel, start, sex, doctor, pain, take, test, thing
Diagnosis (results)	test, positive, negative, sex, chlamydia, result, herpes, partner, day, month
Penile “bump” symptom (no pictures, no herpes)	com, imgur, bump, help, penis, look, know, pimple, hurt, think
Herpes image	imgur, com, herpes, bump, red, penis, help, day, look, month

[a]STD: sexually transmitted disease.

[b]HPV: human papillomavirus.

Table 3. Postvaccination topics and the top 10 terms derived from a Latent Dirichlet Allocation model created from 3 different time periods: pre–COVID-19, prevaccination, and postvaccination.

Topic	Top 10 terms
Herpes symptoms/diagnosis	test, herpes, HSV[a], sex, know, outbreak, negative, positive, genital, risk
Herpes image	com, imgur, herpes, sex, look, help, remove, know, oral, bump
HPV[b]/warts treatment, herpes questioning	wart, herpes, ibb_co, com, www_reddit, comment, remove, HPV, month, skin
Diagnosis/chlamydia	day, month, feel, start, sex, doctor, pain, take, test, thing
Testing, no specific etiology	test, remove, STD, day, sex, week, negative, help, time, oral
Penile “bump” symptom (no pictures, no herpes)	penis, bump, sex, day, know, STD[c], feel, look, condom, time
Oral sex/STD questioning	sex, know, test, week, say, think, time, symptom, oral, tell
Penile “bump” symptom	bump, com, look, imgur, week, ago, day, penis, red, notice

[a]HSV: herpes simplex virus.

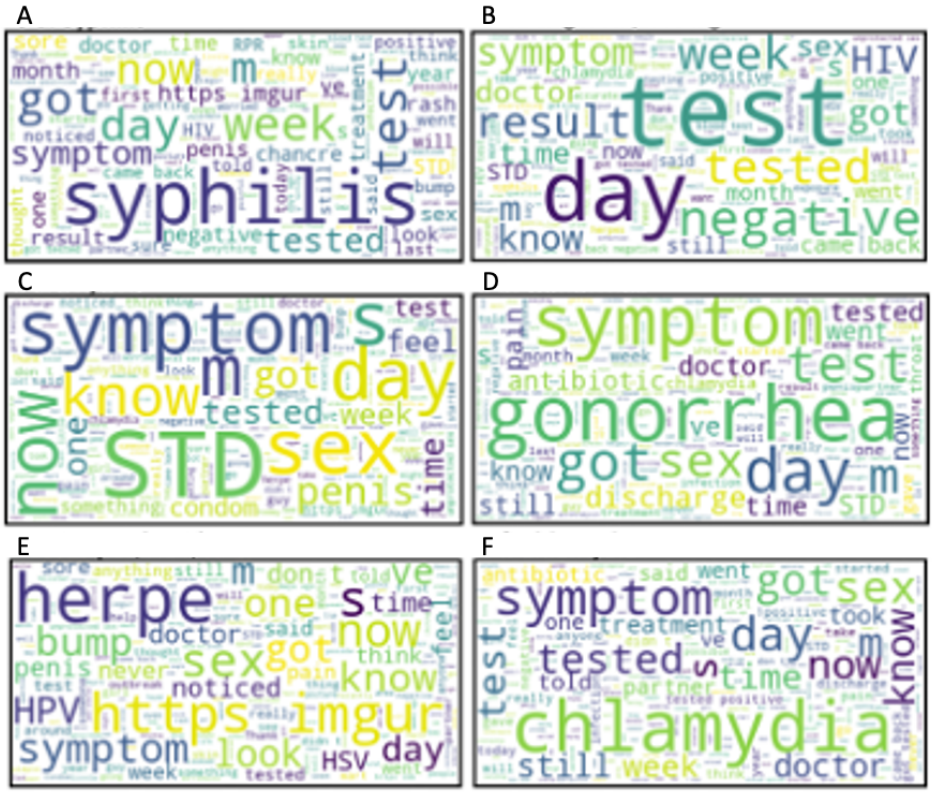
[b]HPV: human papillomavirus.

[c]STD: sexually transmitted disease.

Word Clouds

Although the terms in the topic models listed above are informative, we used *WordCloud* visualizations to better understand the relative importance of these words within each topic based on etiology and general terms over the study period. The terms that appear larger appeared more frequently within the topic, whereas the terms in smaller font appeared less frequently. Figure 6A-F displays the word clouds for 6 specific topics; for example, Figure 6E displays terms clustered with herpes/HSV/HPV such as “imgur” (denoting that a picture was uploaded), “bump,” “pain,” and “outbreak.”

Figure 6. Word clouds by key term: (A) syphilis, (B) diagnosis/testing, (C) STI/STD, (D) gonorrhea, (E) herpes/HSV/HPV, and (F) chlamydia. HPV: human papillomavirus; HSV: herpes simplex virus; STD: sexually transmitted disease; STI: sexually transmitted infection.



Spline Regressions

As shown in the series of spline regressions in Figure 7A-F, there is some variability in the percentage of posts by key terms over the study periods. Although some regressions are "flat" (ie, uniform) across the study periods, others display variability that coincides with the COVID-19 periods. For example, Figure 7F displays the regression for posts with the key terms "diagnose/test/tested." There is some variability in the percentage of posts in the different COVID-19 periods, with the pre–COVID-19 and postvaccination volumes being similar to each other and the prevaccination period having a lower percentage of posts. We tested the significance of the differences between the pre–COVID-19 and post–COVID-19 frequency of posts for 6 key terms (Table 4). There are statistically significant differences for herpes and syphilis, with post–COVID-19 frequencies being higher than pre–COVID-19 frequencies.

Figure 7. Percentage of Reddit posts containing specific key terms from March 2019 to July 2021: (A) gonorrhea, (B) chlamydia, (C) syphilis, (D) gonorrhea/dysuria/discharge, (E) herpes/HSV/HPV, and (F) diagnose/test/tested. HPV: human papillomavirus; HSV: herpes simplex virus.

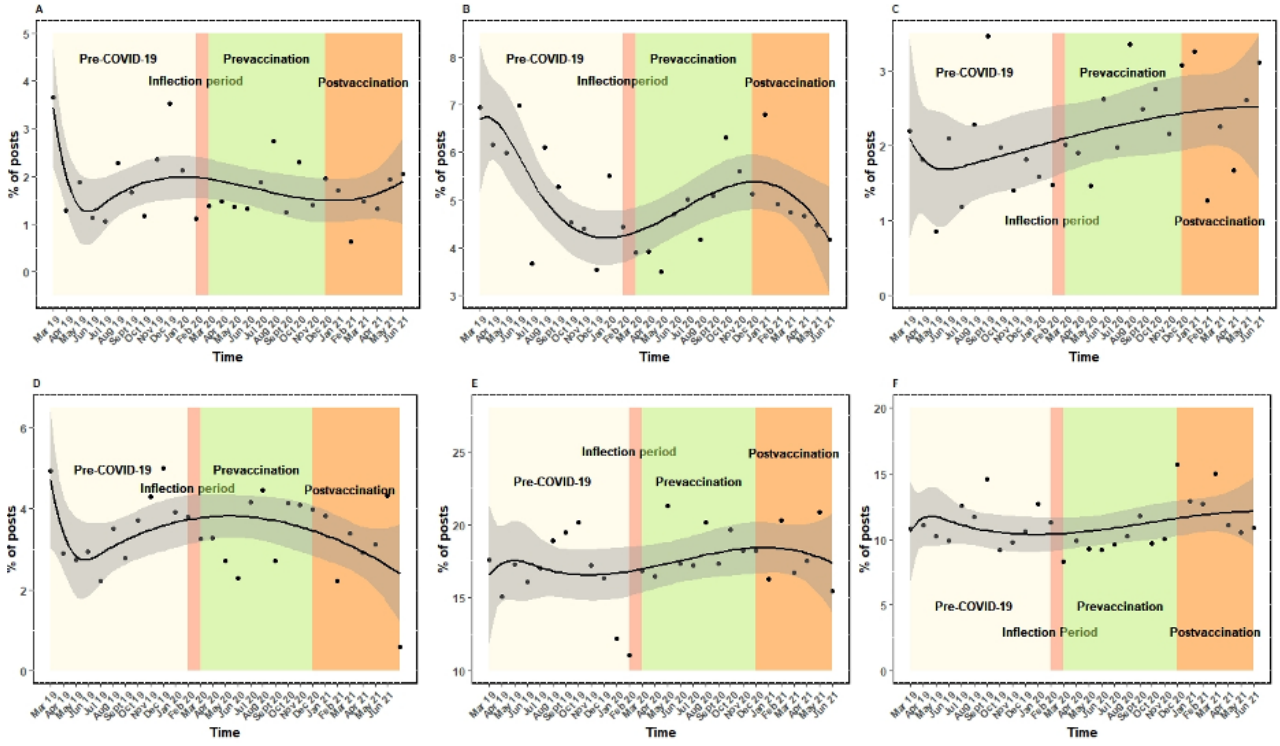


Table 4. Comparing the differences between the pre–COVID-19 and post-COVID-19 frequencies of posts.

Key term	Pre–COVID-19, frequency	Post–COVID-19, frequency	*P* value
Chlamydia	35.27	43.47	.08
Gonorrhea	14.25	15.56	.64
Herpes	118.75	160.47	.01
Syphilis	12.92	21.23	.005
Test/diagnosis	78	97.65	.08
Gonorrhea/dysuria	25.33	30.47	.28

Discussion

Principal Findings

Our study provides evidence that there was an increase in the volume of STD-related posts during the COVID-19 pandemic periods and there were changes in the topics posted in STD-related subreddits from pre–COVID-19 through COVID-19, prevaccination and COVID-19, postvaccination periods. The changes in topics discussed likely relate to behavior changes due to COVID-19–related lockdowns, restrictions on in-person gatherings, and the closing of nonessential medical services [27]. Regardless of lockdown status, people still engage in sexual behavior (eg, condomless sex) that will expose them to STDs. However, with the reduction of STD testing or treatment, these cases are not reflected in surveillance numbers. It is important to understand the sexual health experiences of communities, including symptoms, questions, and behavior patterns, to plan for screening and treatment options.

Our results found that "STD risk" as a topic and general "risk" terms as words only appeared in the pre–COVID-19 time period, whereas "HPV" and "warts" only appeared in the COVID-19, prevaccination and postvaccination periods. During the pre–COVID-19 time period, users generated posts related to general STD risk and sexual behavior, seeking advice and support for understanding STD exposure risk for specific sexual behavior or partnership choices. During the 2 COVID-19 periods, this general "STD risk" topic no longer appeared, demonstrating a difference in content—moving from general discussions to specific symptom or etiology-based posts. During the 2 COVID-19 periods, HPV/warts emerged as a topic. This finding may be due to increased effort to self-diagnosis symptoms experienced as a result of limited access to diagnostic services. Although reported STD cases declined during the initial lockdown period, cases reported in 2020 quickly rebounded and exceed the case numbers in 2019 [3].

Our study's use of Reddit is a novel way to gain insights into STD symptoms experienced, potential exposures, testing decisions, common questions, and behavior patterns (eg, during

lockdown periods). For example, reduction in STD screening may result in observed negative health outcomes due to missed cases, which also impacts onward transmission. The reduction in access to STD testing and treatment during COVID-19 intensified existing barriers to sexual health care, including stigma, judgement, cost, and accessibility [28]. It is important that STD services be maintained, either through telehealth and in-home testing options or via clinic services with COVID-19 mitigation procedures in place (screening, masking, and social distancing).

As Reddit use is anonymous, users may discuss sensitive topics with greater detail and more freely than in clinical encounters. The sexual health subreddits had an average volume of unique posts ranging from approximately 700 to 900 per month; thus, Reddit is a frequently used source of information that could guide the understanding of the behavior, symptoms, and common questions of patients. Sexual health care workers should consider collaboration with Reddit or other social media outlets to leverage the potential benefits of these platforms (anonymous, free, and rapid response) while mitigating harm (incorrect diagnoses and faulty recommendations) [29].

Limitations

Study results should be interpreted while considering the following limitations. LDA is an unsupervised approach with no gold standard to compare to. However, we analyzed the LDA output qualitatively with the use of 2 independent coders and reached 100% consensus on manual topic labels. As we used posts from an open web-based forum, we were unable to validate users; however, there is little incentive to be dishonest or to post false information on health-related subreddits. Reddit users tend to be younger and are more likely to be male compared to the larger US population; however, other demographic trends (eg, race/ethnicity) mirror the distribution in the United States [30]. As men and Black or African American and Latino communities are often underrepresented in STI case data, it is important to gain an understanding of their sexual health needs and experiences via alternative data sources [30]. Finally, the precise location of Reddit users are unknown. Although we were able to extract posts limited to the United States and those in the English language, we cannot pinpoint post volume by specific state or local jurisdiction.

Conclusion

This study demonstrates Reddit as having feasibility and utility to enhance the understanding of sexual behaviors, STD experiences, and needed health engagement with the public. It is important to prioritize efforts to reduce the spread and impact of STDs through surveillance, screening, and treatment. The COVID-19 pandemic and subsequent stay-at-home orders highlight a critical need for increased access to STD clinics and STD information. Data from anonymous Reddit posts may be leveraged to enhance the understanding of the distribution of disease and need for targeted outreach or screening programs.

Conflicts of Interest

None declared.

References

1. Sexually Transmitted Disease Surveillance 2019. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/std/statistics/2019/ [accessed 2022-10-07]
2. Sexually transmitted infections prevalence, incidence, and cost estimates in the United States. Centers for Disease Control and Prevention. 2021 Feb 18. URL: https://www.cdc.gov/std/statistics/prevalence-incidence-cost-2020.htm [accessed 2022-10-07]
3. Pagaoa M, Grey J, Torrone E, Kreisel K, Stenger M, Weinstock H. Trends in nationally notifiable sexually transmitted disease case reports during the US COVID-19 pandemic, January to December 2020. Sex Transm Dis 2021 Oct 01;48(10):798-804 [FREE Full text] [doi: 10.1097/OLQ.0000000000001506] [Medline: 34224523]
4. Bachmann LH, Thorpe P, Bolan G, Mermin J. STD treatment options. Centers for Disease Control and Prevention. 2020 Apr 06. URL: https://www.cdc.gov/nchhstp/dear_colleague/2020/dcl-040620-std-treatment-options.html [accessed 2022-10-07]
5. COVID-19 and the state of the STD field. National Coalition of STD Directors. 2020 May 12. URL: https://www.ncsddc.org/resource/covid-19-and-the-state-of-the-std-field/ [accessed 2022-10-07]
6. Rushmore J, Copen CE, Schneider J, Lamuda PS, Taylor BG, Kirkcaldy RD, et al. Partner seeking and sexual behavior in the United States during the COVID-19 pandemic, March 2020 to March 2021. Sexual Trans Dis 2021 Nov 12;49(4):e57-e60. [doi: 10.1097/olq.0000000000001582]
7. Fox S, Dugan M. Health online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-10-07]
8. Johnson AK, Mikati T, Mehta SD. Examining the themes of STD-related Internet searches to increase specificity of disease forecasting using internet search terms. Sci Rep 2016 Nov 09;6:36503 [FREE Full text] [doi: 10.1038/srep36503] [Medline: 27827386]
9. Anderson KE. Ask me anything: what is Reddit? Library Hi Tech News 2015 Jul 06;32(5):8-11. [doi: 10.1108/lhtn-03-2015-0018]
10. Silberman WR, Record RA. We post it, U Reddit: exploring the potential of reddit for health interventions targeting college populations. J Health Commun 2021 Jun 03;26(6):381-390. [doi: 10.1080/10810730.2021.1949648] [Medline: 34260329]

11. Kurtti A, Cohen M, Jagedo J. Analysis of Reddit reveals dupilumab questions among atopic dermatitis patients. J Drugs Dermatol 2022 Mar 01;21(3):292-294. [doi: 10.36849/JDD.5942] [Medline: 35254758]
12. Pilkington PD, Bedford-Dyer I. Mothers' worries during pregnancy: a content analysis of Reddit posts. J Perinat Educ 2021 Apr 01;30(2):98-107 [FREE Full text] [doi: 10.1891/J-PE-D-20-00018] [Medline: 33897234]
13. Mason A, Jang K, Morley K, Scarf D, Collings SC, Riordan BC. A content analysis of Reddit users' perspectives on reasons for not following through with a suicide attempt. Cyberpsychol Behav Soc Netw 2021 Oct;24(10):642-647. [doi: 10.1089/cyber.2020.0521] [Medline: 33601950]
14. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 01;3:993-1022 [FREE Full text]
15. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. 2020 May 26 Presented at: Fourteenth International AAAI Conference on Web and Social Media (ICWSM-20); June 8-11, 2019; Atlanta, GA p. 830-839 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7347
16. Mohseni MM, Benard RB, Mead-Harvey C, Mi L, Lindor RA, Sheele JM. Sexually transmitted infections in the emergency department are not associated with holidays or school breaks. Am J Emerg Med 2021 Jul;45:642-644. [doi: 10.1016/j.ajem.2020.11.020] [Medline: 33246859]
17. Shah AP, Smolensky MH, Burau KD, Cech IM, Lai D. Recent change in the annual pattern of sexually transmitted diseases in the United States. Chronobiol Int 2007;24(5):947-960. [doi: 10.1080/07420520701648325] [Medline: 17994348]
18. Vijayarani S, Ilamathi J, Nithya S. Preprocessing techniques for text mining - an overview. International Journal of Computer Science & Communication Networks 2015;5(1):7-16 [FREE Full text]
19. Řehůřek R, Sojka P. Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 2011;3(2).
20. Loper E, Bird S. NLTK: the natural language toolkit. In: Association for Computational Linguistics. 2002 Jul 07 Presented at: ETMTNLP '02: ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics; July 7, 2002; Philadelphia, PA p. 63-70. [doi: 10.3115/1118108.1118117]
21. Honnibal M, Montani I. Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Unpublished software application. spaCy. 2017. URL: https://spacy.io [accessed 2022-10-07]
22. Wang C, Blei DM. Collaborative topic modeling for recommending scientific articles. 2011 Aug 21 Presented at: KDD '11: 17th ACM SIGKDD international conference on Knowledge discovery and data mining; August 21-24, 2011; San Diego, CA p. 448-456. [doi: 10.1145/2020408.2020480]
23. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. 2015 Feb 02 Presented at: WSDM '15: Eighth ACM International Conference on Web Search and Data Mining; February 2-6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
24. Mueller A. wordcloud 1.8.2.2. Python Package Index. 2013. URL: https://pypi.python.org/pypi/wordcloud [accessed 2021-11-01] [WebCite Cache ID 6oD8n8hqa]
25. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham, Switzerland: Springer; 2016.
26. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. BMC Med Res Methodol 2019 Mar 06;19(1):46 [FREE Full text] [doi: 10.1186/s12874-019-0666-3] [Medline: 30841848]
27. Nagendra G, Carnevale C, Neu N, Cohall A, Zucker J. The potential impact and availability of sexual health services during the COVID-19 pandemic. Sex Transm Dis 2020 Jul;47(7):434-436 [FREE Full text] [doi: 10.1097/OLQ.0000000000001198] [Medline: 32520878]
28. Tao J, Napoleon SC, Maynard MA, Almonte A, Silva E, Toma E, et al. Impact of the COVID-19 pandemic on sexually transmitted infection clinic visits. Sex Transm Dis 2021 Jan;48(1):e5-e7 [FREE Full text] [doi: 10.1097/OLQ.0000000000001306] [Medline: 33181578]
29. Nobles AL, Leas EC, Althouse BM, Dredze M, Longhurst CA, Smith DM, et al. Requests for diagnoses of sexually transmitted diseases on a social media platform. JAMA 2019 Nov 05;322(17):1712-1713 [FREE Full text] [doi: 10.1001/jama.2019.14390] [Medline: 31688875]
30. Sattelberg W. The demographics Of Reddit: who uses the site? Alphr. 2021 Apr 06. URL: https://www.alphr.com/demographics-reddit/#:~:text=According%20to%20data%20collected%20in,percent%20were%2065%20or%20older [accessed 2021-11-01]

Abbreviations

HPV: human papillomavirus
HSV: herpes simplex virus
LDA: Latent Dirichlet Allocation
STD: sexually transmitted disease
STI: sexually transmitted infection

Edited by C Basch; submitted 12.02.22; peer-reviewed by SF Tsao, A Rovetta; comments to author 31.07.22; revised version received 01.09.22; accepted 20.09.22; published 31.10.22.

Please cite as:
Johnson AK, Bhaumik R, Nandi D, Roy A, Mehta SD
Sexually Transmitted Disease–Related Reddit Posts During the COVID-19 Pandemic: Latent Dirichlet Allocation Analysis
J Med Internet Res 2022;24(10):e37258
URL: https://www.jmir.org/2022/10/e37258
doi:10.2196/37258
PMID:36219757

Original Paper

Empowering Health Care Workers on Social Media to Bolster Trust in Science and Vaccination During the Pandemic: Making IMPACT Using a Place-Based Approach

Shikha Jain[1*], MD; Serena R Dhaon[2*], BS; Shivani Majmudar[3*], MSJ; Laura J Zimmermann[4*], MD, MS; Lisa Mordell[5*], MS; Garth Walker[6*], MD; Amisha Wallia[7*], MD, MPH; Halleh Akbarnia[8*], MD; Ali Khan[9*], MD, MPH; Eve Bloomgarden[10*], MD; Vineet M Arora[11*], MD, MAPP

[1]Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States
[2]Chicago College of Osteopathic Medicine, Midwestern University, Downers Grove, IL, United States
[3]College of Medicine, University of Illinois at Chicago, Chicago, IL, United States
[4]Department of Preventive Medicine, Department of Internal Medicine, Rush University, Chicago, IL, United States
[5]Loyola University Stritch School of Medicine, Maywood, IL, United States
[6]Department of Emergency Medicine, Northwestern University, Evanston, IL, United States
[7]Department of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Northwestern University, Evanston, IL, United States
[8]Advocate Condell Medical Center, Libertyville, IL, United States
[9]Oak Street Health, Chicago, IL, United States
[10]NorthShore University HealthSystem, Evanston, IL, United States
[11]University of Chicago Pritzker School of Medicine, Chicago, IL, United States
[*]all authors contributed equally

Corresponding Author:
Shikha Jain, MD
Department of Medicine
University of Illinois at Chicago
1740 W Taylor St
Chicago, IL, 60612
United States
Phone: 1 773 960 2567
Email: sjain03@gmail.com

Abstract

Background: Given the widespread and concerted efforts to propagate health misinformation on social media, particularly centered around vaccination during the pandemic, many groups of clinicians and scientists were organized on social media to tackle misinformation and promote vaccination, using a national or international lens. Although documenting the impact of such social media efforts, particularly at the community level, can be challenging, a more hyperlocal or "place-based approach" for social media campaigns could be effective in tackling misinformation and improving public health outcomes at a community level.

Objective: We aimed to describe and document the effectiveness of a place-based strategy for a coordinated group of Chicago health care workers on social media to tackle misinformation and improve vaccination rates in the communities they serve.

Methods: The Illinois Medical Professionals Action Collaborative Team (IMPACT) was founded in March 2020 in response to the COVID-19 pandemic, with representatives from major academic teaching hospitals in Chicago (eg, University of Chicago, Northwestern University, University of Illinois, and Rush University) and community-based organizations. Through crowdsourcing on multiple social media platforms (eg, Facebook, Twitter, and Instagram) with a place-based approach, IMPACT engaged grassroots networks of thousands of Illinois health care workers and the public to identify gaps, needs, and viewpoints to improve local health care delivery during the pandemic.

Results: To address vaccine misinformation, IMPACT created 8 "myth debunking" infographics and a "vaccine information series" of 14 infographics that have generated >340,000 impressions and informed the development of vaccine education for the Chicago Public Libraries. IMPACT delivered 13 policy letters focusing on different topics, such as health care worker personal

protective equipment, universal masking, and vaccination, with >4000 health care workers signatures collected through social media and delivered to policy makers; it published over 50 op-eds on COVID-19 topics in high-impact news outlets and contributed to >200 local and national news features. Using the crowdsourcing approach on IMPACT social media channels, IMPACT mobilized health care and lay volunteers to staff >400 vaccine events for >120,000 individuals, many in Chicago's hardest-hit neighborhoods. The group's recommendations have influenced public health awareness campaigns and initiatives, as well as research, advocacy, and policy recommendations, and they have been recognized with local and national awards.

Conclusions: A coordinated group of health care workers on social media, using a hyperlocal place-based approach, can not only work together to address misinformation but also collaborate to boost vaccination rates in their surrounding communities.

(J Med Internet Res 2022;24(10):e38949) doi:10.2196/38949

KEYWORDS

misinformation; COVID-19; place-based; infodemic; infographic; social media; advocacy; infodemiology; vaccination; health care worker; policy maker; health policy; community health

Introduction

With the COVID-19 pandemic came an onslaught of misinformation that undermined the ability to keep the pandemic in check. Because misinformation has been declared a public health crisis by the surgeon general of the United States, individual clinicians are advised to partake in tackling misinformation, particularly when using social media [1]. However, addressing health misinformation can be particularly challenging for individual health care workers who were often serving on the frontlines in the pandemic. Engaging in such activities is even more challenging for health care workers with caregiving responsibilities or those whose communities were disproportionately affected by the pandemic [2].

Health care professionals continue to be trusted sources of information, though fewer Americans trust social media as a source of information [3]. Studies show that social media posts from clinicians are more trusted than posts by political figures [4]. However, health care workers often face attacks and harassment for spreading medical information during the pandemic, particularly due to the increasing polarization of our populace on political lines, with respect to mitigation strategies [5,6].

Although social media has the ability to bring together health care workers to tackle misinformation and promote public health on a national scale in the digital space, there are still major questions that remain unanswered about such efforts. For example, linking social media efforts to actual patient or clinical outcomes continues to be a major challenge, especially on social media, where clinicians are often distributed over large distances.

A specific place-based approach, used across other sectors focused on both the social and physical environment of a community, may be particularly impactful to address misinformation, while simultaneously engaging a local community to improve outcomes [7,8]. During a pandemic, this may be of particular importance, for example, for vaccination rates in a community. To date, few social media efforts describe such a place-based approach to address misinformation, subsequently affecting health processes or outcomes in public health, such as vaccination rates in a community. The aim of this paper is to provide a descriptive process analysis of a multimodal initiative that specifically coupled place-based theory with a social media strategy to combat misinformation related to COVID-19. The goal was to mobilize a virtual health care community to reduce COVID-19 vaccine hesitancy and improve vaccine access in underserved areas served by this community.

Methods

The Illinois Medical Professionals Action Collaborative Team (IMPACT) is an interdisciplinary coalition of health care professionals, founded in March 2020 and established as a 501(c)(3) nonprofit organization in 2022, in response to the COVID-19 pandemic. IMPACT [9] leverages social media and novel partnerships to (1) identify and amplify public health needs and inequities in care delivery, (2) address needs and gaps by rapidly disseminating evidence-based information, (3) connect groups to resources, and (4) advocate for science-based policy.

IMPACT started as an extension of a well-established, closed clinician social media–based group (on Facebook) of >2800 verified physician members in the Chicagoland area. The Facebook group was created in 2015 and is moderated by an IMPACT cofounder and one of the authors (LZ). In March 2020, a post by a clinician in this Facebook group highlighted public health concerns regarding the community spread of COVID-19 and brought attention to the need for preventive measures (eg, social distancing). At the same time, Chicago-based social media physician advocates (VA, SJ, and AK) were sharing similar concerns on another social media platform (Twitter), sharing images of densely packed crowds at both Chicago Saint Patrick's Day celebrations and Chicago's O'Hare International airport [10-12]. As a result, members of the group, now known as IMPACT, mobilized on social media to write a letter that amassed over 300 signatures from verified health care workers (crowdsourced on social media and shared networks) and delivered it to the governor to address the concerns of the health care community [13,14]. The concerns stemmed primarily from lack of social distancing observed in local communities, despite the continued reinforcement by IMPACT members' own health care facilities of the need to distance in the hospitals to avoid being infected. The rapid response strategy of crowdsourcing on Illinois-based health care worker networks on Facebook, followed by an amplification

on Twitter by Chicago-based social media health care advocates and leaders, demonstrated that a coordinated place-based (eg, a local coordinated community) strategy could be effective on social media to influence policy changes and help advise the community during this challenging time. To link these various efforts across social media platforms, IMPACT was formed by team members across disciplines and expertise with social media advocacy work across multiple platforms. Members of the organization were from diverse backgrounds, including community members and organizations, health care workers, and health care organizations. Networks and expertise were combined to create an infrastructure to (1) identify community needs (ie, public health measures), (2) garner information about resource allocation and misallocation, (3) rapidly inform and amplify important trusted information, and (4) alert policy makers and the community of key trusted information or areas of confusion. Later, with the help of trainees (ie, students and residents) and other volunteers, IMPACT expanded the place-based approach and social media strategy to other social media outlets such as Instagram, LinkedIn, and TikTok (@rubin_allergy).

IMPACT immediately recognized that this place-based social media strategy was effective in communicating and amplifying issues that community-based organizations were facing, including personal protective equipment (PPE) procurement and the need for masks. IMPACT subsequently partnered with 2 student-led, community-based organizations—MasksNow Illinois and Get Us PPE Chicago—to help support and fundraise for masking in disadvantaged communities, homeless shelters, and nursing homes. For example, when a need for cloth masks was identified in Little Village—a predominantly Spanish-speaking community in Chicago's West Side—IMPACT leaders connected community leaders with MasksNow Illinois to supply over 15,000 masks to this community. As a result of this early advocacy, a feature story on IMPACT appeared in the Chicago Tribune that focused on the origins of the group, a derivation primarily from physician mothers through the Facebook Chicago group [15]. This piece led to a weekly segment for one of our founders (SJ) during the morning news on Fox32—the local Fox television affiliate. The recurring news segment served as a vehicle to provide information to the community about public health measures [16].

As the COVID-19 pandemic progressed, it was abundantly clear that local communities would benefit greatly from the scientific advancement through COVID-19 vaccines. Unfortunately, it was also recognized that progress would be impeded through inequities in both distribution and uptake in vaccination rates at the local level. IMPACT, therefore, leveraged its social media platforms and strategy to specifically begin to communicate crucial trusted information about vaccines in addition to helping support vaccination events across the Chicagoland area.

IMPACT created infographics to rapidly and effectively communicate key messages to other health professionals and to the general public [17]. Infographics were created by a core "digital media" team of volunteer student interns, a physician lead, and IMPACT cofounders. Infographics drafted by this core team were then reviewed and edited for a lay audience with an eighth-grade literacy. Citations and dates were included and confirmed in each infographic. Infographics were then reviewed by the broader IMPACT team and translated into Spanish by bilingual IMPACT members. Once approved, infographics were shared across social media. As of April 2022, IMPACT designed and shared over 60 infographics.

In addition to social media and traditional media, IMPACT was also called upon to provide grassroots educational efforts with trusted messengers to reach out to the communities most impacted by COVID-19. Since the beginning of the pandemic, over 80 virtual community town halls have taken place in both English and Spanish to inform communities about COVID-19. In addition to these community town halls, IMPACT was contacted by a National Library of Medicine All of Us grantee to generate English and Spanish "train the trainer" videos and training for the Chicago Public Library librarians to serve as trusted messengers to the community on the COVID-19 vaccine topic.

IMPACT additionally expanded and cosponsored large vaccination efforts across Chicago and the greater Chicagoland areas. Many of the vaccination events have occurred in areas of low vaccination status and in partnership with community organizations, places of faith, and schools already trusted by local communities. When Emergency Use Authorization of COVID-19 vaccines for children aged 5-11 years was announced, IMPACT addressed concerns of the communities, and we continue to collaborate with partners to vaccinate and protect children.

The key to the IMPACT strategy has been multistakeholder collaborations. IMPACT additionally partnered with multiple academic institutions in the Chicagoland area to launch key educational initiatives. At the Pritzker School of Medicine at the University of Chicago, it developed and launched a new course on misinformation and science communication for early-year medical students; one of the students' final projects was featured [18] at a national symposium sponsored by the Association of American Medical Colleges [19]. Students at the University of Chicago and the University of Illinois also participated in an op-ed accelerator, which has resulted in over 12 op-eds with student first authors. At the University of Illinois, IMPACT formed a formal internship program with medical students involved in the urban medicine elective to train the next generation of physicians in advocacy.

IMPACT continues to support and develop vaccination events with community partners, create information campaigns across the digital and traditional media space, and pioneer community events with trusted messengers to educate the public and address misinformation surrounding the ongoing pandemic. All of IMPACT's work rests on the volunteer time of health care workers or health professional trainees who do this work in addition to their daytime work and responsibilities.

Results

Our evaluation centers around IMPACT's activities in public policy and the media, social media campaign reach, as well as vaccine education and outreach in the Chicago community. Due

to the nature of this work, we focus on the reach of our social media efforts, using a variety of metrics readily available on social media platforms, as well as both the reach and effectiveness of our social media vaccination efforts.

Policy and Media Reach

IMPACT has delivered 13 policy letters with >4000 health care workers signatures collected through social media and delivered to policy makers and has published >50 op-eds in local and national media (eg, US News, Chicago Tribune, Health Affairs, USA Today, CNN, and Newsweek). IMPACT members have been featured in national media, including the New York Times, Time Magazine, Washington Post, Good Morning America, and National Public Radio and have appeared in >200 local and national news media, educating the public about COVID-19 mitigation and prevention.

Social Media Reach

IMPACT social media campaigns have resulted in Facebook, Twitter, Instagram, and LinkedIn pages with >4000 followers, and posts earning >20,000 views on Facebook alone, as well as a newly verified Twitter account indicated by a blue checkmark. Successful campaigns included a social distancing hashtag (#6ftApartNotUnder) with >4000 tweets and millions of impressions, a universal mask mandate petition on change.org website with >113,000 signatures, a virtual #WhiteCoatsforBlackLives march with >1 million impressions, and COVID-19 data infographics with >400,000 impressions. The number of social media impressions is defined as the number of times content is displayed on a user's screen or within their feed, regardless of whether it is clicked on or interacted with.

Vaccine Education and Outreach

IMPACT created a "myth debunking" series of 8 infographics and a "vaccine information series" of 14 infographics that were shared across social media platforms and used by local schools, health departments, advocacy organizations, and community outreach events [20]. These two infographics series alone generated >330,000 impressions from social media based on an analysis of IMPACT social media metrics. Later, volunteers at community vaccine events were trained to use IMPACT infographics in printed forms as educational resources. Digital and print copies were also provided to community members to help inform them of the importance of vaccination.

To help improve vaccination rates, disparities in vaccination needs were rapidly identified through multiple sources (eg, Twitter, Chicago Facebook health care worker groups, emails, and messages to IMPACT) for health care workers not affiliated with health systems. An IMPACT clearinghouse for vaccine information (eg, registration and interest surveys) was created, procuring information rapidly through social media and professional networks [21]. In response to the concerns IMPACT raised regarding the accessibility of vaccines to those unaffiliated with hospitals, local health departments encouraged all health care entities to vaccinate non–system-affiliated health care workers. This work was highlighted in the Chicago Mayor's weekly press conference with an IMPACT representative [22].

During the later phases of vaccination, when vaccines were available to the general public, IMPACT played a leading role, using social media networks to mobilize >700 health care workers along with >1000 nonmedical volunteers to staff vaccine events in multiple communities [21]. To date, IMPACT has organized or assisted in approximately 400 vaccine events, which have resulted in the vaccination of over 120,000 individuals in the Chicago region. This includes the administration of 5545 pediatric vaccine doses and 6456 booster doses given at 45 different pediatric and booster vaccination events since November 2021.

For this work, IMPACT has received multiple recognitions. The 2021 Community Activism Award from the Democrats of Northfield Township was presented to the group by Governor Pritzker. A community award, the Leadership Legacy Award, was granted by one of the first schools that had approached IMPACT for help. This school benefited from IMPACT's virtual town halls for parents, students, and teachers, followed by actual on-site vaccine events, bringing the percentage of teachers vaccinated from under 50% to 95%. IMPACT was also one of the four recipients of the 2022 Innovations to Bolster Community Trust and Engagement in Science Award from the Association of American Medical Colleges [23]. One of our members (HA) also received the WGN News Remarkable Women Award for her efforts in organizing the vaccine clearinghouse and mobilizing volunteers to vaccinate Chicagoland communities [24].

Discussion

This is the first description of a group of health care professionals using a predominantly web-based strategy on social media to engage in place-based advocacy during the COVID-19 pandemic. This paper reports on the relatively underreported area of health advocacy through social media, an approach with increasing relevance due to an increase in misinformation on social media and the need for health care professionals to address the needs of communities in real-time crises such as the COVID-19 pandemic. Social media crowdsourcing and digital collaborations with local stakeholders and various experts allowed the group to identify and counteract misinformation, identify and amplify the different needs of local communities, and direct both information and resources more equitably.

It is worth considering what is novel about this approach and when and why it might be successful. First, because the COVID-19 pandemic magnified health care disparities, it accelerated the need to develop innovative strategies to provide both information and resources to underserved communities on social media and on the ground. Although the creation of infographics to address misinformation is not new, repurposing infographics created primarily for use in the digital space to be used in physical forms at vaccination events can help "train the trainer" and highlight the role of community members as vehicles to dispel misinformation [25]. The pandemic also laid bare the vulnerabilities of the health care system and the interdependence of hospital capacity, public health messaging, and politics. Health care professionals are trusted voices in the

community with the expertise to counter misinformation and advocate for the needs of their patients. Although individual efforts remain important and have been well described in health care in the past [26], this novel method of collective advocacy with a community or place-based approach has the ability to amplify and leverage evidence-based information on social media and other platforms, with the possibility of timely change for the public in real time. The use of a place-based approach is particularly unique in that, and if used as described, it can help leverage the power of social media engagement and mobilize the local community to fill gaps and address inequities identified at the community level. A place-based approach also lends itself to advocacy efforts, given that much of local policy is settled on a local and state level, particularly during the pandemic, as states and cities faced different community infection rates and differences in community needs. Therefore, such types of advocacy can be used to improve resource alignment and outreach in communities most in need.

There are limitations to this descriptive process analysis. The most significant limitation is the inability to measure the effectiveness of our social media interventions by measuring outcomes among social media users and community members exposed to our work; this requires formal research, and future research may include a randomized controlled trial as well as a qualitative study to further assess the effectiveness of our interventions. Outcomes demonstrating effectiveness include respondent knowledge, attitudes, perceptions, and practices, following exposure to our interventions. Future work may also include testing a more formal process of cocreation of infographics and other materials with members of audience communities. Nonetheless, there is value in the metrics of social media reach and the anecdotal evidence provided; this paper demonstrates a proof of concept and the feasibility of a place-based approach to health care advocacy through social media, laying the groundwork for formal research studies. There are also limitations to IMPACT's entirely volunteer-based model described in this study; although IMPACT is now a 501(c)(3) nonprofit organization, we were initially unable to raise money and relied on the volunteer time of busy clinicians and students to execute this work. It is possible that a model with fundraising would have greater reach and effectiveness compared to our model.

In conclusion, IMPACT describes how a place-based approach on social media across multiple platforms can be repurposed not only to combat misinformation at the community level but also to advocate for science-based policies, engage stakeholders, and help direct resources to organizations and communities most in need. This proof-of-concept application of a social media strategy with a place-based approach may be useful to address other public health needs such as gun violence or the opioid epidemic. Further exploration of such approaches is warranted.

Acknowledgments

We would like to acknowledge the many members of IMPACT (Illinois Medical Professionals Action Collaborative Team) who are not listed in our author team but have contributed to the outcomes and initiatives of this study.

Conflicts of Interest

None declared.

References

1. Murthy VH. U.S. Department of Health & Human Services. 2021 Jul 22. URL: https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf [accessed 2022-03-30]
2. Soares A, Thakker P, Deych E, Jain S, Bhayani RK. The impact of COVID-19 on dual-physician couples: a disproportionate burden on women physicians. J Womens Health (Larchmt) 2021 May;30(5):665-671 [FREE Full text] [doi: 10.1089/jwh.2020.8903] [Medline: 33751922]
3. Solnick RE, Chao G, Ross RD, Kraft-Todd GT, Kocher KE. Emergency physicians and personal narratives improve the perceived effectiveness of COVID-19 public health recommendations on social media: a randomized experiment. Acad Emerg Med 2021 Feb 27;28(2):172-183 [FREE Full text] [doi: 10.1111/acem.14188] [Medline: 33263357]
4. Julie R. Young people rely on social media, but don't trust it. Gallup. 2021 Nov 18. URL: https://news.gallup.com/opinion/gallup/357446/young-people-rely-social-media-don-trust.aspx [accessed 2022-03-30]
5. Arora VM, Bloomgarden E, Jain S. Supporting health care workers to address misinformation on social media. N Engl J Med 2022 May 05;386(18):1683-1685 [FREE Full text] [doi: 10.1056/nejmp2117180]
6. Rubin R. When physicians spread unscientific information about COVID-19. JAMA 2022 Mar 08;327(10):904-906. [doi: 10.1001/jama.2022.1083] [Medline: 35171224]
7. Place-based approaches to supporting children and families. Melbourne Centre for Community Child Health, The Royal Children's Hospital. 2011. URL: https://www.rch.org.au/uploadedFiles/Main/Content/ccch/Policy_Brief_23_-_place-based_approaches_final_web2.pdf [accessed 2022-08-04]
8. Leininger LJ, Albrecht SS, Buttenheim A, Dowd JB, Ritter AZ, Simanek AM, et al. Fight like a nerdy girl: the dear pandemic playbook for combating health misinformation. Am J Health Promot 2022 Mar 14;36(3):563-567. [doi: 10.1177/08901171211070956] [Medline: 35164543]

9. IMPACT: Illinois Medical Professionals Action Collaborative Team. URL: https://www.impact4hc.com/ [accessed 2022-08-04]
10. Hey @PodSaveAmerica @jonlovett @TVietor08 + the real @jonfavs - while on #COVID19, can you help #medtwitter out? Tons of St. Patrick's Day bar crawl madness. We need to convince people to stay home and drink over Slack convos #FlattenTheCurve. @alikhan28. 2020 Apr 14. URL: https://twitter.com/alikhan28/status/1239015803200167937?s=20&t=Q8S4HZXAKI4RncFC2wEaOQ [accessed 2022-08-04]
11. We started putting the letter together and circulating at midnight last night. Thank you #physicianmommiesChicago #IllinoisCOVID19MedicalCollective. @LJoyZ_MD. 2020 Apr 15. URL: https://twitter.com/LJoyZ_MD/status/1239302494884372480?s=20&t=L_PyM0kcyy88KFJ4mzV7ug [accessed 2022-08-04]
12. Huge #Chicago action to #FlattenTheCurve #StayTheFHome #COVID19 thanks to collective action led by @FutureDocs @ShikhaJainMD and countless other health professionals, community advocates + citizens for society #COVID－19 #medtwitter #hcldr. @alikhan28. 2020 Apr 15. URL: https://twitter.com/alikhan28/status/1239286807864451072?s=20&t=Q8S4HZXAKI4RncFC2wEaOQ [accessed 2022-08-04]
13. IMPACT4HC. Social Distancing Letter to Government. URL: https://static1.squarespace.com/static/5e80f9e9abd507573ed7f6c0/t/5e8a09f11b981f2a96bd1f68/1586104817738/SocialDistancingLetterDeliveredtoGovernor3_15_2020.pdf [accessed 2022-03-30]
14. Dear @GovPritzker , Over 300 IL Physicians are calling for urgent action to enforce social distancing given Chicago mass gatherings (e.g. St. Patrick's Day, O'hare). Help us #FlattenTheCurve now. #COVID19 https://docs.google.com/document/d/1uvDDiTm6DsIq_nT3aaQyPOYofR3YgO4VJP8yYopTnRo/edit?usp=sharing cc @alikhan28 @ShikhaJainMD @LJoyZ @haroldpollack. @FutureDocs. 2020 Apr 15. URL: https://twitter.com/FutureDocs/status/1239271918479474688?s=20&t=QHr_1Z0OgBuWdpoCeSU9zg [accessed 2022-08-04]
15. Thayer K. Chicago doctors — mostly moms — join together to advise officials and the public. ‘We can’t not say something.’. Chicago Tribune. 2020 May 07. URL: https://www.chicagotribune.com/coronavirus/ct-life-coronavirus-mom-doctors-impact-group-tt-20200507-vyvtdwkpyzbytp2udja7ikpxyq-story.html [accessed 2022-03-30]
16. Jain S. Chicago doctors come together to dispel misinformation, advise officials on pandemic. Fox32 Chicago. 2020 May 14. URL: https://www.fox32chicago.com/video/684221 [accessed 2022-03-30]
17. Rotolo SM, Jain S, Dhaon S, Dokhanchi JK, Kalata E, Shah T, et al. A coordinated strategy to develop and distribute infographics addressing COVID-19 vaccine hesitancy and misinformation. J Am Pharm Assoc (2003) 2022;62(1):224-231 [FREE Full text] [doi: 10.1016/j.japh.2021.08.016] [Medline: 34483058]
18. Tesema N, Collison M, Luo C. Using medical students as champions against misinformation during a global pandemic. AAMC 2022 Aug 01;97(8):1103-1104. [doi: 10.1097/ACM.0000000000004368] [Medline: 34406130]
19. Pritzker SOM. Pritzker students take on misinformation. The University of Chicago Pritzker School of Medicine. 2021 Oct 27. URL: https://pritzker.uchicago.edu/news/pritzker-students-take-misinformation [accessed 2022-08-04]
20. IMPACT infographics. IMPACT. URL: https://www.impact4hc.com/impact-infographics [accessed 2022-03-30]
21. Wallia A, Khan A, Zimmermann LJ, Desai A, Blankemeier J, Bloomgarden E, et al. Creation and translation of impact (Illinois medical professional action collaborative team) to amplify and address disparities in covid-19: the case of vaccine delivery. Health Serv Res 2021 Sep 15;56(S2):33-33. [doi: 10.1111/1475-6773.13764]
22. "If we are going to get to maximal impact, we must move fast and slow" - @alikhan28 @OakStreetHealth *We have a lot of work to do for non-hospital affiliated healthcare workers. *We must also address vaccine hesitancy and do the hard work. #protectchicago #VaxUpIL. @IMPACT4HC. 2021 Jan 12. URL: https://twitter.com/IMPACT4HC/status/1349074204130488321?s=20&t=iRnXHkamEMaDslatGDPGo [accessed 2022-08-04]
23. 2022 AAMC innovations that bolster community trust in science award. Association of American Medical Colleges. URL: https://www.aamc.org/what-we-do/aamc-awards/innovations-in-research [accessed 2022-03-30]
24. Congrats to our remarkable women winner!. WGN 9. 2022 Apr 04. URL: https://wgntv.com/news/remarkable-women/congrats-to-our-remarkable-women-winner/ [accessed 2022-08-04]
25. Becker SJ, Helseth SA, Tavares TL, Squires DD, Clark MA, Zeithaml VA, et al. User-informed marketing versus standard description to drive demand for evidence-based therapy: a randomized controlled trial. Am Psychol 2020 Nov;75(8):1038-1051 [FREE Full text] [doi: 10.1037/amp0000635] [Medline: 33252943]
26. Cobey JC. The doctor-activist: physicians fighting for social change. JAMA 1996 Nov 27;276(20):1688. [doi: 10.1001/jama.1996.03540200074040]

Abbreviations

IMPACT: Illinois Medical Professionals Action Collaborative Team
PPE: personal protective equipment

Edited by M Gisondi, M Gottlieb; submitted 22.04.22; peer-reviewed by S Gordon, T Turk; comments to author 21.06.22; revised version received 08.07.22; accepted 13.07.22; published 17.10.22.

Please cite as:
Jain S, Dhaon SR, Majmudar S, Zimmermann LJ, Mordell L, Walker G, Wallia A, Akbarnia H, Khan A, Bloomgarden E, Arora VM Empowering Health Care Workers on Social Media to Bolster Trust in Science and Vaccination During the Pandemic: Making IMPACT Using a Place-Based Approach
J Med Internet Res 2022;24(10):e38949
URL: https://www.jmir.org/2022/10/e38949
doi:10.2196/38949
PMID:35917489

Original Paper

Efficacy, Usability, and Acceptability of a Chatbot for Promoting COVID-19 Vaccination in Unvaccinated or Booster-Hesitant Young Adults: Pre-Post Pilot Study

Tzu Tsun Luk[1], PhD; Judy Hiu Tung Lui[1]; Man Ping Wang[1], PhD

School of Nursing, The University of Hong Kong, Pokfulam, China (Hong Kong)

Corresponding Author:
Man Ping Wang, PhD
School of Nursing
The University of Hong Kong
3 Sassoon Road
Pokfulam
China (Hong Kong)
Phone: 852 39176636
Email: mpwang@hku.hk

Abstract

Background: COVID-19 vaccines are highly effective in preventing severe disease and death but are underused. Interventions to address COVID-19 vaccine hesitancy are paramount to reducing the burden of COVID-19.

Objective: We aimed to evaluate the preliminary efficacy, usability, and acceptability of a chatbot for promoting COVID-19 vaccination and examine the factors associated with COVID-19 vaccine hesitancy.

Methods: In November 2021, we conducted a pre-post pilot study to evaluate "Vac Chat, Fact Check," a web-based chatbot for promoting COVID-19 vaccination. We conducted a web-based survey (N=290) on COVID-19 vaccination at a university in Hong Kong. A subset of 46 participants who were either unvaccinated (n=22) or were vaccinated but hesitant to receive boosters (n=24) were selected and given access to the chatbot for a 7-day trial period. The chatbot provided information about COVID-19 vaccination (eg, efficacy and common side effects), debunked common myths about the vaccine, and included a decision aid for selecting vaccine platforms (inactivated and mRNA vaccines). The main efficacy outcome was changes in the COVID-19 Vaccine Hesitancy Scale (VHS) score (range 9-45) from preintervention (web-based survey) to postintervention (immediately posttrial). Other efficacy outcomes included changes in intention to vaccinate or receive boosters and willingness to encourage others to vaccinate on a scale from 1 (not at all) to 5 (very). Usability was assessed by the System Usability Scale (range 0-100). Linear regression was used to examine the factors associated with COVID-19 VHS scores in all survey respondents.

Results: The mean (SD) age of all survey respondents was 21.4 (6.3) years, and 61% (177/290) of respondents were female. Higher eHealth literacy (B=–0.26; *P*<.001) and perceived danger of COVID-19 (B=–0.17; *P*=.009) were associated with lower COVID-19 vaccine hesitancy, adjusting for age, sex, chronic disease status, previous flu vaccination, and perceived susceptibility to COVID-19. The main efficacy outcome of COVID-19 VHS score significantly decreased from 28.6 (preintervention) to 24.5 (postintervention), with a mean difference of –4.2 (*P*<.001) and an effect size (Cohen *d*) of 0.94. The intention to vaccinate increased from 3.0 to 3.9 (*P*<.001) in unvaccinated participants, whereas the intention to receive boosters increased from 1.9 to 2.8 (*P*<.001) in booster-hesitant participants. Willingness to encourage others to vaccinate increased from 2.7 to 3.0 (*P*=.04). At postintervention, the median (IQR) System Usability Scale score was 72.5 (65-77.5), whereas the median (IQR) recommendation score was 7 (6-8) on a scale from 0 to 10. In a post hoc 4-month follow-up, 82% (18/22) of initially unvaccinated participants reported having received the COVID-19 vaccine, whereas 29% (7/24) of booster-hesitant participants received boosters.

Conclusions: This pilot study provided initial evidence to support the efficacy, usability, and acceptability of a chatbot for promoting COVID-19 vaccination in young adults who were unvaccinated or booster-hesitant.

(J Med Internet Res 2022;24(10):e39063) doi:10.2196/39063

KEYWORDS

COVID-19; coronavirus; vaccine; immunization; booster; vaccine hesitancy; chatbot; conversational agent; virtual assistant; Chinese; young adult; youth; health promotion; health intervention; chatbot usability; pandemic; booster hesitancy; web-based survey; students; university students

Introduction

COVID-19 vaccines are highly effective in preventing severe disease and death but are underused. By mid-2022, the full vaccination rate has remained suboptimal in many places where COVID-19 vaccines are readily available (eg, 67% in the United States and 75% in the United Kingdom) [1]. COVID-19 booster shots are also being delivered to address waning immunity and viral variants, but studies have shown that some fully vaccinated people were unwilling to take the booster [2-4]. COVID-19 may also become an endemic disease such as seasonal influenza, and regular vaccination may be needed to protect high-risk populations. Interventions to promote the vaccine is crucial to reduce the burden of COVID-19.

Vaccine hesitancy is considered 1 of the 10 major threats of global health according to the World Health Organization (WHO) [5]. Studies have consistently shown higher COVID-19 vaccine hesitancy in women, younger people, ethnic minority populations, and people with lower socioeconomic status [6,7]. Partly due to the fast-tracked development and authorization of the vaccines, the lack of confidence in the vaccine efficacy and safety were among the main drivers for hesitancy [7]. Widespread misinformation against the vaccine further amplified its safety concerns [8,9]. Debunking such misinformation could reduce COVID-19 vaccine hesitancy and promote uptake, especially in subpopulations that are more susceptible to misinformation such as young people [10].

Chatbots or conversational agents are increasingly developed as a scalable and accessible platform for supporting health care delivery. The interface of a chatbot that is familiar to most people with experiences in mobile messaging could promote the usability and user engagement of the chatbot compared to other platforms. Several chatbots have been developed amid the COVID-19 pandemic [11,12], mostly for symptom checking and information dissemination [13,14]. The WHO has also launched chatbots on popular social networking sites such as WhatsApp to provide instant and credible information about COVID-19, including vaccination [15]. Nevertheless, empirical evidence on the utility of chatbots for promoting vaccination has remained scarce.

Mass COVID-19 immunization has begun in February 2021 in Hong Kong. Despite ample supply of both an inactivated vaccine (CoronaVac; Sinovac) and mRNA vaccine (Comirnaty; Fosun-BioNTech), the uptake had been slowed, with only 62% of the population being fully vaccinated by the beginning of 2022 [1]. We conducted a population-based survey on 1501 general adults in Hong Kong (COVID-19 Health Information Survey) and found higher COVID-19 vaccine hesitancy among young adults (aged 18-29 years) than older adults (aged ≥30 years) [16,17]. We also found the low perceived COVID-19 severity and safety concerns of the vaccines to be the main drivers of vaccine hesitancy [16]. Additionally, the COVID-19 Health Information Survey showed that eHealth literacy was associated with adherence to mask wearing, hand washing and social distancing [17], but its role in vaccine hesitancy has remained under-studied. Therefore, the primary aim of the study was to examine the preliminary efficacy, usability, and acceptability of using a chatbot for promoting COVID-19 vaccination. We also examined the feasibility of assessing the long-term effect on COVID-19 vaccination status in a post-hoc 4-month follow-up. The secondary aim was to examine the factors associated with COVID-19 vaccine hesitancy, including eHealth literacy.

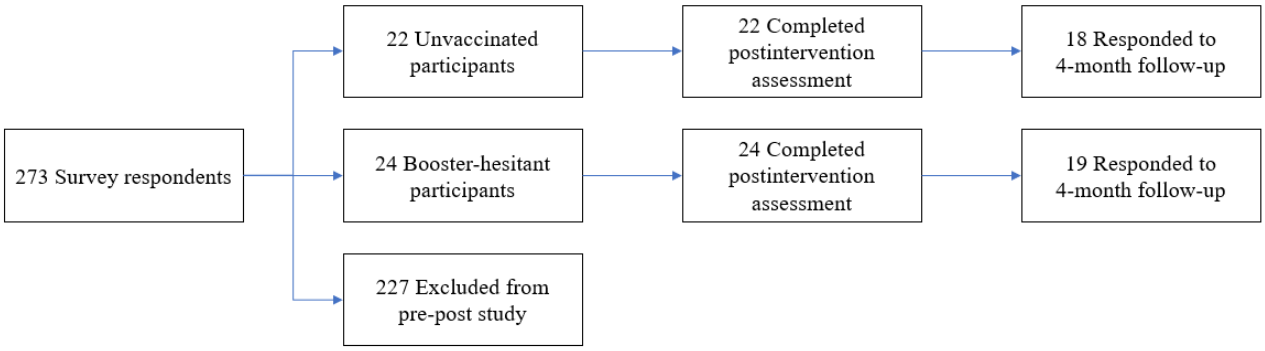
Methods

Study Design and Recruitment

We conducted a pilot study using a pretest-posttest design to evaluate “Vac Chat, Fact Check,” a chatbot designed to provide updated information and debunk misinformation about the COVID-19 vaccine. The study was conducted in November 2021, between the end of the fourth wave (June 2021) and the start of the fifth wave (January 2022) of the outbreak in Hong Kong, which had about 12,000 cumulative cases and 200 deaths.

The study targeted adults aged ≥18 years who can read and communicate in Chinese. A mass email with a link to a web-based survey of COVID-19 vaccination was sent to all students at a public university in Hong Kong on November 8, 2021. The survey link was open for 7 days and received 290 valid responses. Of these, 273 (94.1%) respondents indicated their interest in participating in the pre-post evaluation of the chatbot by leaving their contact information at the end of the survey. We identified and invited all 46 respondents who were either unvaccinated (n=22) or fully vaccinated but hesitant to receive boosters if eligible (n=24; response rate: 46/46, 100%). The planned sample size (20-25 each for unvaccinated or booster-hesitant participants) was based on a previous formative study of a chatbot for promoting human papillomavirus vaccination [18]. Figure 1 shows the study flow diagram.

Figure 1. Study flow diagram.



Ethics Approval

The study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 21-449).

Study Procedures

Participants who were invited to participate in the pre-post study received a WhatsApp message describing the study purpose and provided informed consent. The participants then received a URL link to access the web-based chatbot and start a 7-day trial period. The chatbot could be accessed repeatedly. WhatsApp reminders to use the chatbot were sent on day 3, day 5, and day 7. On day 8, we sent a URL link to the postintervention questionnaire. Participants who completed the pre-post study were given HK $300 (US $38.5) for their time and effort.

On March 30, 2022, about 4 months after the completion of the pre-post study, we conducted a post hoc follow-up with a single question about COVID-19 vaccination or booster status via WhatsApp. Participation was voluntary, and consent was obtained from those who responded to the question. The additional follow-up served to examine the feasibility of measuring the long-term effect of the chatbot.

Design of the Chatbot

The “Vac Chat, Fact Check” chatbot was developed by our team. To promote dissemination, the chatbot could be accessed by any internet browser on smartphones, tablets, and personal computers (ie, web-based). The chatbot was available in Chinese only since most Hong Kong residents (>90%) spoke Chinese. Upon entering the chatbot, the user received a message about how to use the chatbot and a menu of options showing the core functionality of the chatbot. The users could navigate the chatbot by typing the corresponding number of the options in the menus or keywords (eg, allergy) to obtain information directly (Figures 2 and 3).

The chatbot conversation generally unfolded by following predefined rules or decision trees. To better simulate human interactions, the chatbot also used natural language processing (NLP) powered by Google Dialogflow for handling small talk (eg, greetings and thank-yous). The chatbot provided responses mainly using texts with emojis, but some messages also included infographics (Figure 3).

The intervention content followed the Confidence, Complacency and Convenience (“3C’s”) model of vaccine hesitancy [19]. Specifically, the information addressed the lack of trust in the effectiveness and safety of the vaccine (confidence), the lack of perceived risk of COVID-19 or the perception that vaccination is not necessary (complacency), and barriers to access the vaccine (convenience). The information provided by the chatbot was categorized into 6 major topics (Table 1). Our population-based survey suggested that inadequate knowledge about COVID-19 could contribute to vaccine hesitancy [16]. Therefore, the chatbot included general information about COVID-19. Since 2 types of vaccine (inactivated and mRNA vaccines) were available in Hong Kong with differing eligibility criteria (age and pregnancy status), the chatbot also included a decision aid for selecting the suitable vaccine.

Figure 2. Screenshot of "Vac Chat, Fact Check" showing chatbot navigation by menu options.

English translation of the messages:

You can enter a number to select the following information or service or enter your question directly:

Select the service you need:
1. Information on COVID-19
2. Information on COVID-19 vaccine
3. Why should I receive the vaccine?
4. Am I suitable for vaccination?
5. Myths and facts about COVID-19 vaccines
6. Information on boosters
7. Book now

Enter the number for details:

1. Eligibility for boosters
2. Why should I receive a booster?
3. Advice on booster vaccination

Enter 0 to return to the main menu

The Government announced that, from 1 January 1, 2022, onward:

1. ...

Figure 3. Screenshot of "Vac Chat, Fact Check" showing chatbot navigation by keyword.

English translation of the messages and infographics:

Allergy

Are people with allergies suitable for COVID-19 vaccination?

Most people with allergies can receive COVID-19 vaccines!

Consult a doctor if you have 1 of the following 4 conditions:
1. Anaphylaxis (severe allergic reaction)
2. Allergic reaction (within 1 hour) to multiple drugs
3. Suspected allergic reaction to a previous COVID-19 vaccine
4. Allergy to vaccines, polyethylene glycol (PEG), or polysorbate

The chance of having a severe allergic reaction to the COVID-19 vaccine is extremely low. People with mild allergies are suitable to get vaccinated.

Remark: Severe allergic reaction refers to conditions that require adrenaline or EpiPen© treatment or immediate hospitalization; these do not include asthma, rhinitis, eczema, or hives.

Are people with a history of allergy not suitable for COVID-19 vaccinations?

Facts:
!! Most people with a history of allergy can receive COVID-19 vaccines!

Asthma, rhinitis, and eczema are not contraindications to vaccination. You can receive the vaccine if your condition is stable.
...

Table 1. Overview of the topics and contents of "Vac Chat, Fact Check."

Topic	Content
Information about COVID-19	• Symptoms and complications, including "long COVID" • Route of transmission and incubation period • High-risk populations
Information about COVID-19 vaccination	• Mechanism of the vaccines • Vaccine efficacy • Possible side effects • Eligibility for vaccination
Reasons for getting vaccinated	• Protection of self • Protection of others
Myths and facts about the COVID-19 vaccine	• Alleged side effects (eg, infertility and miscarriage) • Safety of the vaccine (eg, alteration of a person's DNA) • Safety in people with preexisting conditions (eg, a history of allergy) • Lack of efficacy
Information about COVID-19 vaccine boosters	• Eligibility for receiving boosters • Reasons for receiving boosters
Information about how to get vaccinated	• Government's web-based booking system • Venues for vaccination

Instruments

COVID-19 Vaccine–Related Outcomes

All COVID-19 vaccination outcomes were measured at preintervention and postintervention. COVID-19 vaccination status was assessed by asking "have you been vaccinated against COVID-19?" with responses options of "yes, 2 doses," "yes, 1 dose," and "no." Intention to receive the COVID-19 vaccine (for those responded "no") or COVID-19 boosters (for those responded "yes, 2 doses") were assessed on a scale from 1 (not likely at all) to 5 (very likely) [20]. By adapting an item from the OCEANS study [21], we also asked, "if people around you were thinking of getting a COVID-19 vaccination, you would..." Responses were coded from 1 (suggest that they do not get the vaccination) to 5 (strongly encourage them).

The main efficacy outcome was changes in COVID-19 vaccine hesitancy from preintervention to postintervention. We adapted the Vaccine Hesitancy Scale (VHS) developed by the WHO's Strategic Advisory Group of Experts on Immunization for assessing COVID-19 vaccine hesitancy [22]. The COVID-19 VHS included 9 Likert-style items, each coded from 1 (strongly disagree) to 5 (strongly agree; Multimedia Appendix 1 [22,23]). After the reverse coding of some items, all items were summed to give a total score ranging from 9 to 45, with higher scores indicating greater COVID-19 vaccine hesitancy. The VHS can also be divided into the "Lack of confidence" subscale (7 items) and the "Risk" (2 items) subscale and analyzed.

In our sample, the COVID-19 VHS had high internal consistency preintervention (Cronbach α=.86) and postintervention (Cronbach α=.88) [24]. Concurrent validity was supported by a higher mean VHS score in unvaccinated participants versus those who received 1 dose and 2 doses of vaccine (28.6 vs 26.4 vs 23.0, respectively; P<.001). The VHS score was also inversely and moderately correlated to intention to receive the vaccine (Spearman ρ=–0.48; P=.01) or boosters (Spearman ρ=–0.55; P<.001) and willingness to encourage others to vaccinate (Spearman ρ=–0.64; P<.001) [25].

In the 4-month follow-up, we assessed the COVID-19 vaccination status in initially unvaccinated participants by their responses of "no" and "yes, [number] dose(s)." For booster-hesitant participants, we asked whether they had received a booster shot ("yes" or "no").

Usability and Acceptability Outcomes

The postintervention questionnaire included the System Usability Scale (SUS), a widely used instrument in software engineering, to measure the participants' perceived usability of the chatbot. The 10-item SUS gives a composite score ranging from 0 to 100, with 68 or above indicating above-average usability [26]. Other acceptability measures included the perceived usefulness of the chatbot in (1) getting information about the COVID-19 vaccine, (2) making decisions about vaccination, and (3) increasing the motivation to get vaccinated, each assessed on a scale from 1 (not useful at all) to 5 (very useful). Overall satisfaction with the chatbot was assessed by asking "how likely would you recommend the chatbot to other people" on a 11-point scale from 0 (not likely at all) to 10 (very likely).

Other Measures

The baseline questionnaire included the eHealth Literacy Scale (eHEALS) [27], which has been translated into Chinese and used in our study population [28]. The scale included 8 items, which are summed to give an overall score from 8 to 40. Higher scores indicate greater perceived ability to use health technologies. The eHEALS had high internal consistency in our sample (Cronbach α=.91). To assess the perceived susceptibility to and severity of COVID-19, we also asked, "how likely do you think you will contract COVID-19 in the

future?" and "how dangerous do you think COVID-19 is to health?" respectively, each with 11-point response options. Data on sociodemographic characteristics, chronic disease, and previous flu vaccination were also collected.

Statistical Analysis

To evaluate the chatbot efficacy, we used 1-sample, 2-tailed *t* test and Wilcoxon sign-rank test to examine the change in intention to receive a vaccine or booster and COVID-19 VHS scores from preintervention to postintervention. We also examined changes in the "Lack of confidence" and "Risk" subscales of the COVID-19 VHS. The effect size of the pre-post difference in COVID-19 VHS scores (Cohen *d*) was calculated as a mean difference divided by the SD of the mean difference. Our sample size of 46 participants could detect a moderate effect size of 0.43 (Cohen *d*) in the pre-post difference in COVID-19 VHS scores with 80% power at 2-sided 5% level of significance. The corresponding effect sizes detectable were 0.64 for intention to vaccinate (n=21) and 0.60 for intention to receive boosters (n=24). Intervention usability and acceptability were reported descriptively. For the secondary aim, we used bivariate and multivariable linear regression to examine the factors associated with the COVID-19 VHS score in all survey respondents. Factors examined included sociodemographic characteristics, chronic disease status, previous flu vaccination, eHealth literacy, and the perceived susceptibility to and severity of COVID-19.

All statistical analyses were conducted in Stata/MP software (version 15.1; StataCorp). We used complete case analyses because there were no missing data in the web-based survey and postintervention assessment except eHealth literacy (n=2) and the perceived susceptibility (n=4) and severity (n=4) of COVID-19. A 2-sided $P<.05$ denoted statistical significance.

Results

Participant Characteristics

The mean (SD) age of all survey respondents was 21.4 (6.3) years, and 61% (177/290) of respondents were female (Table 2). Participants of the pre-post study (n=46) had similar characteristics to those of nonparticipants (n=244) except, as expected, having significantly higher COVID-19 vaccine hesitancy ($P<.001$).

Table 2. Characteristics of all survey respondents (N=290).

Characteristic	Survey respondents (N=290)	Included in the pre-post study		*P* value[a]
		No (n=244)	Yes (n=46)	
Age (years)				
Mean (SD)	21.4 (6.4)	21.7 (6.8)	20.2 (2.7)	.15
Median (IQR)	20 (19-21)	20 (19-21)	20 (18-21)	.14
Sex, n (%)				
Male	113 (39)	97 (39.8)	16 (35)	.53
Female	177 (61)	147 (60.2)	30 (65)	
Chronic disease, n (%)				
No	270 (93.1)	229 (93.9)	41 (89)	.25
Yes	20 (6.9)	15 (6.1)	5 (11)	
Previous flu vaccination, n (%)				
No	156 (53.8)	134 (54.9)	22 (48)	.38
Yes	134 (46.2)	110 (45.1)	24 (52)	
eHealth literacy[b]				
Mean (SD)	30.2 (4.6)	30.2 (4.7)	30.1 (4.0)	.86
Median (IQR)	32 (28-32)	32 (28-32)	31.5 (28-32)	.56
Perceived susceptibility to COVID-19[c]				
Mean (SD)	3.2 (1.9)	3.2 (1.9)	3.2 (1.6)	.94
Median (IQR)	3 (2-5)	3 (2-5)	3 (2-5)	.68
Perceived severity of COVID-19[c]				
Mean (SD)	6.3 (2.2)	6.3 (2.2)	6.5 (2.2)	.58
Median (IQR)	7 (5-8)	7 (5-8)	7 (5-8)	.61
COVID-19 vaccine hesitancy[d]				
Mean (SD)	23.6 (5.8)	22.6 (5.3)	28.6 (5.6)	<.001
Median (IQR)	23 (20-27)	23 (19-26)	29 (23-33)	<.001

[a]Calculated by chi-squared test, 2-sample, 2-tailed *t* test, or Wilcoxon rank-sum test as appropriate.

[b]Assessed by the eHealth Literacy Scale; possible scores range from 8 to 40, with higher scores indicating greater eHealth literacy.

[c]Assessed by an 11-point scale from 0 to 10; higher scores indicate greater perceived susceptibility or severity.

[d]Assessed by the COVID-19 Vaccine Hesitancy Scale; possible scores range from 9 to 45, with higher scores indicating greater vaccine hesitancy.

Factors Associated With COVID-19 Vaccine Hesitancy

In all survey respondents, both bivariate and multivariable models showed that lower eHealth literacy and perceived danger of COVID-19 were associated with higher COVID-19 vaccine hesitancy (Table 3). The results were similar after additionally adjusting for COVID-19 vaccination status (data not shown).

Table 3. Factors associated with COVID-19 vaccine hesitancy[a] (N=290).

Factor	Crude B (95% CI)	*P* value	Adjusted B (95% CI)[b]	*P* value
Age (years)	0.038 (–0.068 to 0.14)	.48	–0.031 (–0.15 to 0.086)	.60
Sex, female	1.16 (–0.21 to 2.53)	.10	0.79 (–0.061 to 2.18)	.27
Had chronic disease	1.70 (–0.94 to 4.34)	.21	1.65 (–1.24 to 4.54)	.26
Had previous flu vaccination	–1.65 (–5.21 to 1.90)	.35	–0.040 (–1.38 to 1.30)	.95
eHealth literacy[c]	–0.27 (–0.42 to –0.13)	<.001	–0.26 (–0.41 to –0.11)	<.001
Perceived susceptibility to COVID-19[d]	0.17 (–0.20 to 0.54)	.36	0.20 (–0.17 to 0.57)	.29
Perceived severity of COVID-19[d]	–0.35 (–0.66 to –0.050)	.02	–0.41 (–0.71 to –0.10)	.009

[a]Assessed by the COVID-19 Vaccine Hesitancy Scale; possible scores range from 9 to 45, with higher scores indicating greater vaccine hesitancy.

[b]Adjusting for other variables in the table.

[c]Assessed by the eHealth Literacy Scale; possible scores range from 8 to 40, with higher scores indicating greater eHealth literacy.

[d]Assessed by an 11-point scale from 0 to 10; higher scores indicate greater perceived susceptibility or severity.

Pre-Post Evaluation of the Chatbot

The completion rate of the postintervention assessment was 100% (46/46). Table 4 shows the favorable changes in all measures related to COVID-19 vaccination from preintervention to postintervention (mean duration: 15.0 days). The main efficacy outcome of COVID-19 VHS score significantly decreased from 28.6 (preintervention) to 24.5 (postintervention), with a mean difference of –4.2 (P<.001) and an effect size (Cohen d) of 0.94. Similarly, both the "Lack of confidence" and "Risk" subscale scores significantly decreased. Intention to vaccinate or receive boosters and willingness to encourage others to vaccinate significantly increased from preintervention to postintervention. One unvaccinated participant at preintervention reported having received the first dose of the vaccine at postintervention.

Table 4. Changes in COVID-19 vaccine-related measures from preintervention to postintervention (n=46).

	Preintervention	Postintervention	*P* value[a]
COVID-19 vaccine hesitancy (n=46)[b]			
Mean (SD)	28.6 (5.6)	24.5 (6.0)	<.001
Median (IQR)	29 (23-33)	25 (20-29)	<.001
COVID-19 vaccine hesitancy: Lack of confidence (n=46)[c]			
Mean (SD)	20.8 (5.0)	17.2 (5.2)	<.001
Median (IQR)	21 (16-26)	18 (13-21)	<.001
COVID-19 vaccine hesitancy: Risk (n=46)[d]			
Mean (SD)	7.8 (1.3)	7.2 (1.6)	.01
Median (IQR)	8 (7-8)	7.5 (6-8)	.02
Intention to vaccinate (n=21)[e]			
Mean (SD)	3.0 (0.73)	3.9 (0.83)	<.001
Median (IQR)	3 (3-4)	4 (3-4)	.001
Intention to receive boosters (n=24)[e]			
Mean (SD)	1.9 (0.3)	2.8 (0.9)	<.001
Median (IQR)	2 (2-2)	3 (2-3)	<.001
Willingness to encourage others to vaccinate (n=46)[f]			
Mean (SD)	2.7 (1.0)	3.0 (0.9)	.04
Median (IQR)	3 (2-3)	3 (2-4)	.04

[a]Calculated by paired 2-tailed *t* test or Wilcoxon signed-rank test as appropriate.

[b]Assessed by the COVID-19 Vaccine Hesitancy Scale (VHS); possible scores range from 9 to 45, with higher scores indicating greater vaccine hesitancy.

[c]"Lack of confidence" subscale of the COVID-19 VHS; possible scores range from 7 to 35, with higher scores indicating greater lack of confidence in the vaccine.

[d]"Risk" subscale of the COVID-19 VHS; possible scores range from 2 to 10, with higher scores indicating greater perceived risk of the vaccine.

[e]Assessed on a scale from 1 (not likely at all) to 5 (very likely).

[f]Assessed on a scale from 1 (suggest that they do not get the vaccination) to 5 (strongly encourage them).

Usability and Acceptability of the Chatbot

On average, the participants used the chatbot for a total of 64 (SD 47) minutes during the 1-week trial period. Longer time spent on the chatbot was correlated with a larger reduction in vaccine hesitancy with marginal significance (Spearman ρ=0.26; *P*=.08). Among participants who used the chatbot (n=46), the median (IQR) SUS score was 72.5 (65-77.5) out of 100. On a scale from 1 (not agree at all) to 5 (strongly agree), the median (IQR) score on the perceived usefulness of the chatbot was 4 (4-4) for getting information about the COVID-19 vaccine, 3 (2-4) for making decisions about vaccination, and 3 (2-3) for increasing the motivation to get vaccinated. The median (IQR) recommendation score was 7 (6-8) on a scale from 0 to 10.

Vaccination Status at 4-Month Follow-up

Overall, 18 (82%) of the 22 initially unvaccinated participants and 19 (79%) of the 24 booster-hesitant participants responded to the post hoc 4-month follow-up. All 18 unvaccinated participants reported having received COVID-19 vaccination (2 doses: n=16, 89%; and 1 dose: n=2, 11%), whereas 7 (37%) of the 19 booster-hesitant participants reported having received boosters.

Discussion

Principal Findings

This pilot study showed a significant decrease in COVID-19 vaccine hesitancy after using the "Vac Chat, Fact Check" chatbot in young adults who were hesitant to vaccinate or receive boosters. According to the rule of thumb of Cohen [29], the effect size (Cohen *d*=0.94) was large. Other efficacy outcomes, including intention to vaccinate or receive boosters and willingness to encourage others to vaccinate, consistently showed the benefit of the chatbot. The usability of the chatbot was supported by the median SUS score of 72.5 out of 100, which fell between the cutoffs of "good" (a score of 71.4) and "excellent" (a score of 85.5) adjective ratings [30]. The median recommendation score of 7 on a scale from 0 to 10 indicated the satisfactory acceptability of the chatbot [31].

Our PubMed search using the keywords of vaccine and chatbot and their synonyms only identified 1 peer-reviewed study that provided empirical evidence on the efficacy of a chatbot for promoting COVID-19 vaccination. The study was a web-based experiment on a French sample population, which found that

interacting with a chatbot could promote more positive attitudes toward COVID-19 vaccines and intention to vaccinate [32]. A study (preprint) also showed an increase in vaccine acceptance in Japanese adults after using "Corowa-kun," a chatbot in LINE instant messenger [33]. Direct comparison between our study with these studies were difficult because of differences in the study methods, sample characteristics, and outcome measures. Nevertheless, our findings were consistent with these studies by showing a positive impact of chatbot on COVID-19 vaccine uptake.

To our knowledge, our study was the first to include actual receipt of COVID-19 vaccines or boosters as outcome measures in chatbot evaluation. Assuming (conservatively) that all participants lost to follow-up did not receive any vaccine or booster, 82% (18/22) of the initially unvaccinated participants received at least 1 dose of vaccine, whereas 29% (7/24) of booster-hesitant participants received a booster. As a reference, the corresponding rates were 92% and 28% in Hong Kong residents aged 20 to 29 years on March 30, 2022 (same date as the follow-up) [34]. Note that these figures could not be directly compared because of differences in sample characteristics, and our participants were likely more vaccine- or booster-hesitant than the general population. Nonetheless, the satisfactory response rate of 80% (37/46) provides support for the feasibility of conducting longer-term (>3 months) follow-up in future trials.

Corroborating our previous findings in the general population [16], we found that higher perceived severity of, but not susceptibility to, COVID-19 was associated with lower COVID-19 vaccine hesitancy. Previously studies have found that eHealth literacy was associated with knowledge and adherence to nonpharmacological preventive measures against COVID-19 [17,35]. This study further found eHealth literacy to be associated with COVID-19 vaccine hesitancy. Higher eHealth literacy helps people process and discern the credibility of web-based health information, which may buffer the impact of the infodemic (an overabundance of information, both accurate or otherwise, during a disease outbreak) and misinformation against the vaccine and thus hesitancy. Our findings corroborate the importance of building eHealth literacy to fight the COVID-19 pandemic.

Similar to most chatbots built to support health care amid the COVID-19 pandemic [11,12], our "Vac Chat, Fact Check" chatbot was primarily rule-based. We decided against building a chatbot that is entirely powered by NLP for practical reasons. First, existing and readily available NLP engines remain inadequate in handling free-flow conversations in Cantonese (the local Chinese dialect). Second, rule-based chatbots are relatively inexpensive and could be quickly developed and deployed to mitigate the pandemic when health care resources are stretched. NLP-based chatbots could better simulate human interaction but require extensive training and resources to become adequately usable. Nonetheless, our study has provided proof-of-concept evidence to support chatbots as a mode of delivery to promote vaccination, which provides the impetus for developing more sophisticated and potentially more effective chatbots.

Limitations

The main limitation of the pre-post study is the lack of a control or comparison group, which limited the causal inference of any changes observed after using the chatbot. The possibility that the observed changes were attributable to contextual changes along the course of the outbreak could not be excluded. However, the study was conducted at a time when Hong Kong had been maintaining a low level of local transmission with nearly 0 daily local case (from June to December 2021). This setting, coupled with the short interval between preintervention and postintervention assessments, was unlikely to have had a substantial effect on the vaccination outcomes. Nevertheless, the findings must be considered preliminary and hypothesis-generating. Another limitation is the small sample size, which precludes the examination of the chatbot's efficacy in sociodemographic subgroups (eg, sex). Third, since all measures were self-reported, social desirability bias could not be excluded. Finally, our study targeted young adults given their greater vaccine hesitancy than older populations and their frequent use of social networking sites—a major source of misinformation. The generalizability of the findings to other populations is unclear. Due to the convenience sampling method, our participants may not be representative of all young adults who are unvaccinated or booster-hesitant.

Conclusions

Promoting the uptake of COVID-19 vaccines is crucial to mitigating the impact of COVID-19. This pilot study provided initial evidence to support the efficacy, usability, and acceptability of a chatbot for promoting COVID-19 vaccination in young adults who were unvaccinated or booster-hesitant. Randomized controlled trials are warranted to test the effectiveness of the chatbot in increasing COVID-19 vaccination. Although our study indicated the benefits of the chatbot in both unvaccinated and booster-hesitant young adults, the drivers for vaccine hesitancy between the 2 groups likely differ. Further research is also needed to understand their differences to provide more tailored information and optimize the chatbot's efficacy.

Acknowledgments

The study was funded by the Seed Fund for Basic Research, The University of Hong Kong (202011159200). The funder has no involvement in reviewing and approving the manuscript for publication and no role in the conceptualization and implementation of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COVID-19 Vaccine Hesitancy Scale.
[PDF File (Adobe PDF File), 103 KB - jmir_v24i10e39063_app1.pdf]

References

1. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E. Coronavirus (COVID-19) vaccinations. Our World in Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2022-06-30]
2. Rzymski P, Poniedziałek B, Fal A. Willingness to receive the booster COVID-19 vaccine dose in Poland. Vaccines (Basel) 2021 Nov 05;9(11):1286 [FREE Full text] [doi: 10.3390/vaccines9111286] [Medline: 34835217]
3. Pal S, Shekhar R, Kottewar S, Upadhyay S, Singh M, Pathak D, et al. COVID-19 vaccine hesitancy and attitude toward booster doses among US healthcare workers. Vaccines (Basel) 2021 Nov 19;9(11):1358 [FREE Full text] [doi: 10.3390/vaccines9111358] [Medline: 34835289]
4. Paul E, Fancourt D. Predictors of uncertainty and unwillingness to receive the COVID-19 booster vaccine: an observational study of 22,139 fully vaccinated adults in the UK. Lancet Reg Health Eur 2022 Mar;14:100317 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100317] [Medline: 35132400]
5. Ten threats to global health in 2019. World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2022-03-11]
6. Robinson E, Jones A, Lesser I, Daly M. International estimates of intended uptake and refusal of COVID-19 vaccines: a rapid systematic review and meta-analysis of large nationally representative samples. Vaccine 2021 Apr 08;39(15):2024-2034 [FREE Full text] [doi: 10.1016/j.vaccine.2021.02.005] [Medline: 33722411]
7. Cascini F, Pantovic A, Al-Ajlouni Y, Failla G, Ricciardi W. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: a systematic review. EClinicalMedicine 2021 Oct;40:101113 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101113] [Medline: 34490416]
8. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
9. Abbasi J. Widespread misinformation about infertility continues to create COVID-19 vaccine hesitancy. JAMA 2022 Mar 15;327(11):1013-1015. [doi: 10.1001/jama.2022.2404] [Medline: 35191947]
10. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]
11. Almalki M, Azeez F. Health chatbots for fighting COVID-19: a scoping review. Acta Inform Med 2020 Dec;28(4):241-247 [FREE Full text] [doi: 10.5455/aim.2020.28.241-247] [Medline: 33627924]
12. Amiri P, Karahanna E. Chatbot use cases in the COVID-19 public health response. J Am Med Inform Assoc 2022 Apr 13;29(5):1000-1010 [FREE Full text] [doi: 10.1093/jamia/ocac014] [Medline: 35137107]
13. Herriman M, Meer E, Rosin R, Lee V, Washington V, Volpp KG. Asked and answered: building a chatbot to address COVID-19-related concerns. NEJM Catal Innov Care Deliv 2020 Jun 18 [FREE Full text] [doi: 10.1056/CAT.20.0230]
14. Siedlikowski S, Noël LP, Moynihan SA, Robin M. Chloe for COVID-19: evolution of an intelligent conversational agent to address infodemic management needs during the COVID-19 pandemic. J Med Internet Res 2021 Sep 21;23(9):e27283 [FREE Full text] [doi: 10.2196/27283] [Medline: 34375299]
15. WHO Health Alert brings COVID-19 facts to billions via WhatsApp. World Health Organization. 2021 Apr 26. URL: https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp [accessed 2022-03-11]
16. Luk TT, Zhao S, Wu Y, Wong JY, Wang MP, Lam TH. Prevalence and determinants of SARS-CoV-2 vaccine hesitancy in Hong Kong: a population-based survey. Vaccine 2021 Jun 16;39(27):3602-3607 [FREE Full text] [doi: 10.1016/j.vaccine.2021.05.036] [Medline: 34034950]
17. Guo Z, Zhao SZ, Guo N, Wu Y, Weng X, Wong JY, et al. Socioeconomic disparities in eHealth literacy and preventive behaviors during the COVID-19 pandemic in Hong Kong: cross-sectional study. J Med Internet Res 2021 Apr 14;23(4):e24577 [FREE Full text] [doi: 10.2196/24577] [Medline: 33784240]
18. Amith M, Zhu A, Cunningham R, Lin R, Savas L, Shay L, et al. Early usability assessment of a conversational agent for HPV vaccination. Stud Health Technol Inform 2019;257:17-23 [FREE Full text] [Medline: 30741166]
19. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
20. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
21. Freeman D, Loe BS, Yu LM, Freeman J, Chadwick A, Vaccari C, et al. Effects of different types of written vaccination information on COVID-19 vaccine hesitancy in the UK (OCEANS-III): a single-blind, parallel-group, randomised controlled

trial. Lancet Public Health 2021 Jun;6(6):e416-e427 [FREE Full text] [doi: 10.1016/S2468-2667(21)00096-7] [Medline: 33991482]
22. Shapiro GK, Tatar O, Dube E, Amsel R, Knauper B, Naz A, et al. The vaccine hesitancy scale: psychometric properties and validation. Vaccine 2018 Jan 29;36(5):660-667. [doi: 10.1016/j.vaccine.2017.12.043] [Medline: 29289384]
23. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, SAGE Working Group on Vaccine Hesitancy. Measuring vaccine hesitancy: the development of a survey tool. Vaccine 2015 Aug 14;33(34):4165-4175 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.037] [Medline: 25896384]
24. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997 Mar 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
25. Sedgwick P. Spearman's rank correlation coefficient. BMJ 2014 Nov 28;349:g7327. [doi: 10.1136/bmj.g7327] [Medline: 25432873]
26. Brooke J. SUS: a retrospective. J Usability Stud 2013 Feb;8(2):29-40 [FREE Full text]
27. 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]
28. Wong DK, Cheung M. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
29. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic; 1988.
30. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123 [FREE Full text]
31. Krol MW, de Boer D, Delnoij DM, Rademakers JJDJM. The Net Promoter Score--an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109 [FREE Full text] [doi: 10.1111/hex.12297] [Medline: 25345554]
32. Altay S, Hacquin AS, Chevallier C, Mercier H. Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. J Exp Psychol Appl 2021 Oct 28. [doi: 10.1037/xap0000400] [Medline: 34726454]
33. Kobayashi T, Nishina Y, Tomoi H, Harada K, Tanaka K, Matsumoto E. Corowa-kun: impact of a COVID-19 vaccine information chatbot on vaccine hesitancy, Japan 2021. medRxiv. Preprint posted online on May 30, 2021 2022. [doi: 10.1101/2021.05.26.21257854]
34. Statistics on 5th wave of COVID-19 (from 31 Dec 2021 up till 30 Mar 2022 00:00). The Government of Hong Kong Special Administrative Region. URL: https://www.coronavirus.gov.hk/eng/5th-wave-statistics.html [accessed 2022-04-15]
35. An L, Bacon E, Hawley S, Yang P, Russell D, Huffman S, et al. Relationship between coronavirus-related eHealth literacy and COVID-19 knowledge, attitudes, and practices among US adults: web-based survey study. J Med Internet Res 2021 Mar 29;23(3):e25042 [FREE Full text] [doi: 10.2196/25042] [Medline: 33626015]

Abbreviations

eHEALS: eHealth Literacy Scale
NLP: natural language processing
SUS: System Usability Scale
VHS: Vaccine Hesitancy Scale
WHO: World Health Organization

Edited by T Leung, H Gouda, V Arnold; submitted 26.04.22; peer-reviewed by T Kobayashi, Z Rosberger, C Luo, O Tatar, O Rahaman; comments to author 06.06.22; revised version received 01.07.22; accepted 23.09.22; published 04.10.22.

Please cite as:
Luk TT, Lui JHT, Wang MP
Efficacy, Usability, and Acceptability of a Chatbot for Promoting COVID-19 Vaccination in Unvaccinated or Booster-Hesitant Young Adults: Pre-Post Pilot Study
J Med Internet Res 2022;24(10):e39063
URL: https://www.jmir.org/2022/10/e39063
doi:10.2196/39063
PMID:36179132

Original Paper

Using a Chatbot as an Alternative Approach for In-Person Toothbrushing Training During the COVID-19 Pandemic: Comparative Study

Samerchit Pithpornchaiyakul[1,2], DDS, MSc; Supawadee Naorungroj[3], DDS, MS, PhD; Kittiwara Pupong[4], DDS, MS; Jaranya Hunsrisakhun[1,2], DDS, PhD

[1]Department of Preventive Dentistry, Faculty of Dentistry, Prince of Songkla University, Hatyai, Songkhla, Thailand

[2]Improvement of Oral Health Care Research Unit, Faculty of Dentistry, Prince of Songkla University, Hatyai, Songkhla, Thailand

[3]Department of Conservative Dentistry, Faculty of Dentistry, Prince of Songkla University, Hatyai, Songkhla, Thailand

[4]Maelan Hospital, Maelan, Pattani, Thailand

Corresponding Author:
Jaranya Hunsrisakhun, DDS, PhD
Department of Preventive Dentistry
Faculty of Dentistry
Prince of Songkla University
15 Kanjanavanich Rd
Hatyai, Songkhla, 90112
Thailand
Phone: 66 74429875
Email: hjaranya@hotmail.com

Abstract

Background: It is recommended that caregivers receive oral health education and in-person training to improve toothbrushing for young children. To strengthen oral health education before COVID-19, the 21-Day FunDee chatbot with in-person toothbrushing training for caregivers was used. During the pandemic, practical experience was difficult to implement. Therefore, the 30-Day FunDee chatbot was created to extend the coverage of chatbots from 21 days to 30 days by incorporating more videos on toothbrushing demonstrations and dialogue. This was a secondary data comparison of 2 chatbots in similar rural areas of Pattani province: Maikan district (Study I) and Maelan district (Study II).

Objective: This study aimed to evaluate the effectiveness and usability of 2 chatbots, 21-Day FunDee (Study I) and 30-Day FunDee (Study II), based on the protection motivation theory (PMT). This study explored the feasibility of using the 30-Day FunDee chatbot to increase toothbrushing behaviors for caregivers in oral hygiene care for children aged 6 months to 36 months without in-person training during the COVID-19 pandemic.

Methods: A pre-post design was used in both studies. The effectiveness was evaluated among caregivers in terms of oral hygiene practices, knowledge, and oral health care perceptions based on PMT. In Study I, participants received in-person training and a 21-day chatbot course during October 2018 to February 2019. In Study II, participants received only daily chatbot programming for 30 days during December 2021 to February 2022. Data were gathered at baseline of each study and at 30 days and 60 days after the start of Study I and Study II, respectively. After completing their interventions, the chatbot's usability was assessed using open-ended questions. Study I evaluated the plaque score, whereas Study II included an in-depth interview. The 2 studies were compared to determine the feasibility of using the 30-Day FunDee chatbot as an alternative to in-person training.

Results: There were 71 pairs of participants: 37 in Study I and 34 in Study II. Both chatbots significantly improved overall knowledge (Study I: $P<.001$; Study II: $P=.001$), overall oral health care perceptions based on PMT (Study I: $P<.001$; Study II: $P<.001$), and toothbrushing for children by caregivers (Study I: $P=.02$; Study II: $P=.04$). Only Study I had statistically significant differences in toothbrushing at least twice a day ($P=.002$) and perceived vulnerability ($P=.003$). The highest overall chatbot satisfaction was 9.2 (SD 0.9) in Study I and 8.6 (SD 1.2) in Study II. In Study I, plaque levels differed significantly ($P<.001$).

Conclusions: This was the first study using a chatbot in oral health education. We established the effectiveness and usability of 2 chatbot programs for promoting oral hygiene care of young children by caregivers. The 30-Day FunDee chatbot showed the possibility of improving toothbrushing skills without requiring in-person training.

Trial Registration: Thai Clinical Trials Registry TCTR20191223005; http://www.thaiclinicaltrials.org/show/TCTR20191223005 and TCTR20210927004; https://www.thaiclinicaltrials.org/show/TCTR20210927004

(J Med Internet Res 2022;24(10):e39218) doi:10.2196/39218

KEYWORDS

mHealth; tele-dentistry; digital health; chatbot; conversional agents; oral hygiene; oral health behaviors; protection motivation theory; young children; caregiver; in-person toothbrushing training; COVID-19

Introduction

Early childhood caries (ECC) continues to be a significant public health problem worldwide, including in Thailand. ECC can have a negative impact on children's quality of life [1-3] and cost societies and families [4]. Caries is a multifactorial disease in which oral hygiene is a crucial risk factor for developing ECC [5-8]. For young children, having caregivers clean their teeth twice daily with fluoride toothpaste is recommended to prevent ECC [9]. In Thailand, ECC and oral hygiene care for children younger than 3 years remain unsatisfactory. The national oral health survey in 2017 reported a prevalence of dental caries in children aged 3 years of 52.9% and an average number of caries of 2.8 teeth per person [10]. The approach to improving caregivers' toothbrushing behavior with young children has emphasized oral health education in conjunction with in-person training for caregivers [11-14]. However, this personalized approach necessitates time and human resources.

In recent decades, chatbots have been introduced as a new way to improve person-centered health care, particularly to enhance the efficiency of delivery of primary health care services such as health education and counseling support [15,16], which has been shown to increase access to and the quality of services and health information while using fewer human resources [17,18]. Chatbots are computer programs that mimic human conversations using text or voice messages. The technology could be a set of rule-based algorithms or machine learning techniques such as natural language processing to automate some parts of the dialogue [19]. Chatbots are used in a wide range of health care settings [17]; however, they are rare in dentistry.

Zhang et al [20] proposed an artificial intelligence (AI) chatbot behavior change model as a theoretical framework for guiding the design and evaluation of chatbots. This model consists of the following 4 primary components: (1) defining the qualities of the chatbot and gaining an understanding of the user's context, (2) constructing relational ability, (3) building persuasive conversational capacity such as using behavioral change models, and (4) evaluating methods and outcomes. Among the theories of behavioral change for creating persuasive conversational capacity proposed by Zhang et al [20], the protection motivation theory (PMT) [21,22] has been widely accepted [23-25]. PMT is explained by the interplay between threat and coping appraisals, which results in protective health behavior [23]. Threat appraisal combines perceived severity (perceptions of the extent of harm) and perceived vulnerability (perceptions of the likelihood of experiencing harm). Coping appraisal comprises response efficacy (confidence in the effectiveness of the advice in reducing or preventing potential damage) and self-efficacy (belief in one's ability to carry out the recommended behavior successfully), minus response costs (the perceived or actualized costs associated with the practice of the recommended behavior) [21,22]. PMT has been demonstrated to be applicable for behavior change and to have a positive effect on adaptive intentions or behaviors in the varied communication approaches [11,24-26] for which the theory has been suggested for individual and community interventions. Numerous oral health studies have applied and assessed protection motivation variables to determine behavioral intention to self-protective action [11,27], but none have utilized chatbots as an intervention for protective behavior change in dentistry. Therefore, this study intended to use the AI chatbot behavior change model as a framework to enhance the chatbot's usability, satisfaction, and design outcome measures, while PMT was specifically applied to design the content and ways to modify toothbrushing behavior.

A chatbot-based mobile application, a novel PMT-based solution to prevent ECC in Thailand, was launched through the WowBot project in 2018 before the COVID-19 epidemic. The “21-Day FunDee” chatbot and in-person toothbrushing training were implemented in 6 study centers, 5 in rural areas and 1 in an urban area. However, during COVID-19, caregivers with children had difficulty accessing hospitals and other health care service providers, making in-person training impossible. As a result, a modified version entitled “30-Day FunDee” was created to overcome these obstacles, reduce the burden of oral health promotion on health care workers, and decrease the risk of COVID-19 exposure. More video demonstrations and communication focused on in-person practice of toothbrushing techniques and increased motivation for caregivers to improve oral health care for their young children were included in this version. Both chatbots have already been evaluated in sociodemographically comparable populations over different time periods.

Therefore, this secondary data analysis aimed to evaluate the effectiveness and usability of the chatbots before and during the COVID-19 pandemic. In 2 study settings, we demonstrated what we learned about the application of chatbots to promote caregivers’ oral health care practices for young children and the feasibility of using chatbots to improve toothbrushing abilities without actual practice.

Methods

Study Location and Phases

This study performed secondary data analysis of data from the WowBot project, which was conducted between 2018 and 2024. We chose 2 study settings in Pattani province (Study I in Maikan

district and Study II in Maelan district) with comparable socioeconomic backgrounds to illustrate how innovative chatbots encourage caregivers to provide oral health care for their young children in 2 situations: before and during the COVID-19 pandemic.

The WowBot project consisted of 3 phases. Phase I included a scoping review and the development of a chatbot based on PMT. Phase II comprised an assessment of the usability and effectiveness of the first chatbot called "21-Day FunDee" together with in-person toothbrushing training in 6 study centers (Pattani, Phangha, Trang, Songkhla, Nakhonsrithammarat, and Patthalung provinces). Toward the end of phase II, during the COVID-19 pandemic, the chatbot "21-Day FunDee" content was modified with the addition of the following video clips: toothbrushing technique, an examination of dental plaque, and child behavioral management during toothbrushing. The effectiveness of "30-Day FunDee" was evaluated in another community (Maelan District, Pattani province) during the lockdown measures in Thailand. Phase III, an ongoing study, aims to improve chatbot performance and conduct a long-term evaluation of clinical outcomes.

Ethics Approval

The 2 studies were registered with the Thai Clinical Trials Registry and approved by the Faculty of Dentistry, Prince of Songkla University Institutional Review Board (EC6208-031 and EC6407-053).

Chatbot Characteristics and Development

Regarding the design of the chatbots, we applied all 4 components from the model by Zhang et al [20]. For example, in component 1, when designing chatbot characteristics and understanding users' backgrounds, we designed a young female doctor with a friendly, cheerful, and compassionate personality. For component 2, building relational capacity, our chatbot sent caregivers daily funny greeting cards, discussed their challenges, offered emotional support, and concluded with an image or infographic containing an inspirational quote. For component 3, building persuasive conversational capacity, we delivered daily content containing (1) explicit, understandable oral health care knowledge and (2) PMT-based oral health behavior improvement. Last, we applied the fourth component, which was to evaluate mechanisms and outcomes by asking open-ended questions about satisfaction, usage patterns, and how people brushed their teeth (Figures 1 and 2).

The rule-based strategy was used to construct the chatbot flow of "21-Day FunDee" and its modified version, "30-Day FunDee," and the chatbots were developed utilizing the Chatfuel platform. PMT constructs, including perceived severity, perceived vulnerability, response efficacy, and self-efficacy, were used to guide the development of the chatbot content. The chatbots operate on Facebook messenger over the course of 21 and 30 daily sessions, respectively. Through various interactive conversation flows, the chatbots were designed to engage and motivate the users with interesting conversation with rule-based agents. Each session was to run around 3 minutes to 5 minutes and comprises text, video clips, or infographics. The first developed chatbot application, "21-Day FunDee," was reviewed by 2 dentists specializing in community health, while the modified version, "30-Day FunDee," was reviewed by 2 pedodontists. Experts determined the validity of the content based on its accuracy, relevance, and conversational flow. We recruited 10 volunteers to test each chatbot in the following aspects: conversational flow, onboarding, understanding, navigation, response time, and the chatbot's personality.

Figure 1. Examples of user interfaces of the Facebook Messenger–based (A) 21-Day FunDee and (B) 30-Day FunDee including 3 essential elements: greeting; content; and summary containing games, infographics, videos, and friendly conversations.

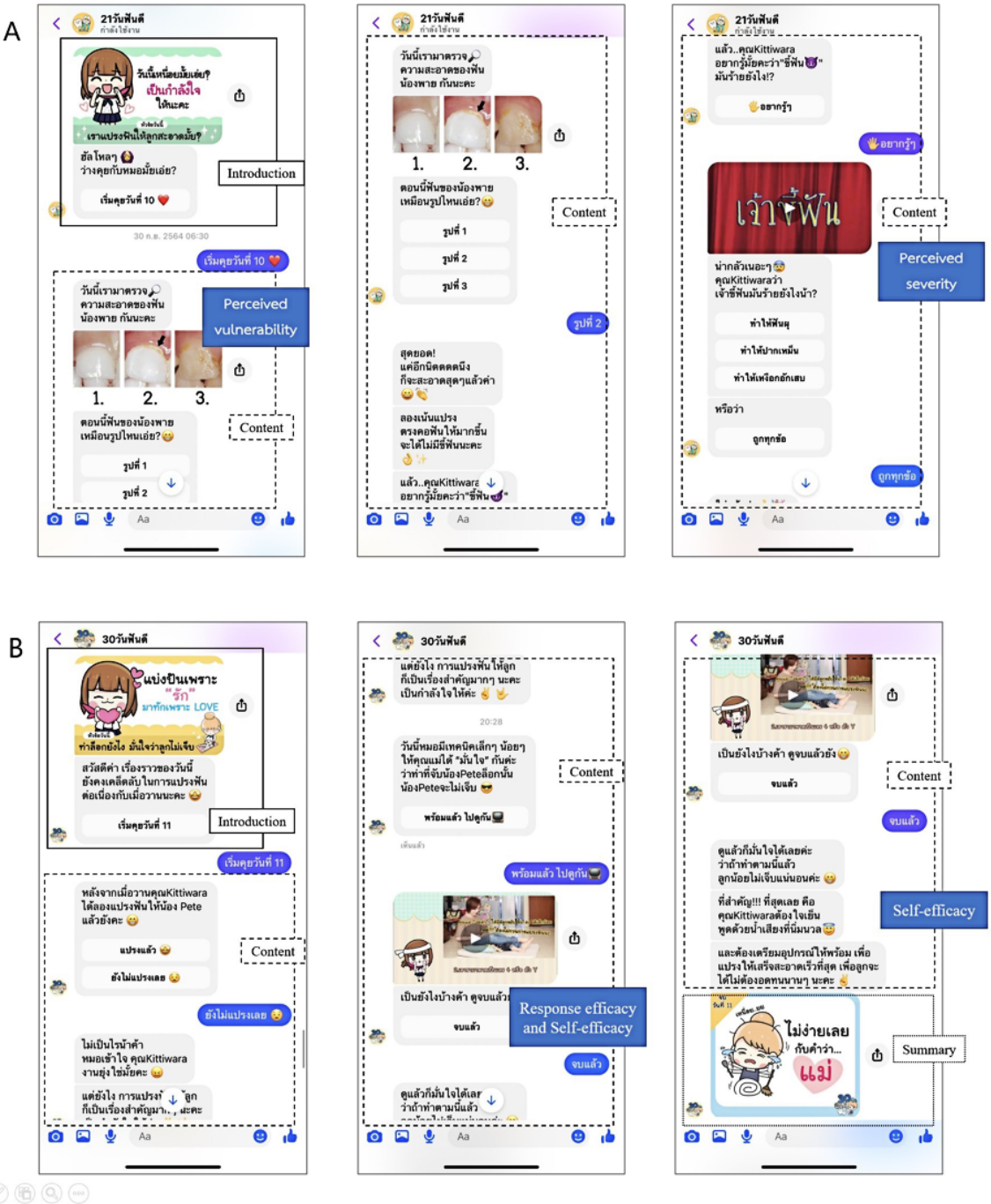
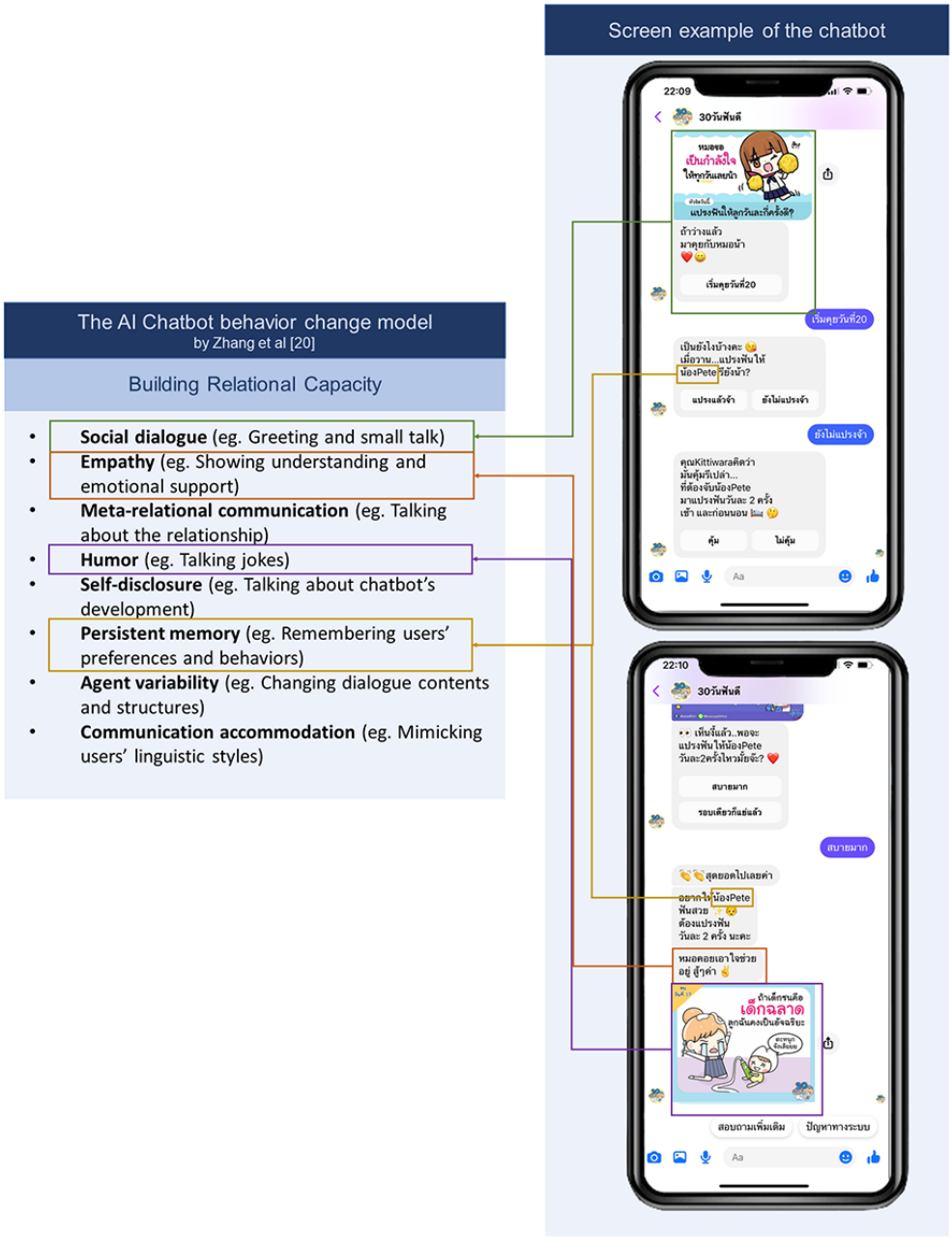


Figure 2. Screenshots showing examples of using the artificial intelligence (AI) chatbot behavior change model.



Setting and Participants

A quasi-experimental design (pretest and posttest) was used to evaluate the chatbots' effectiveness and usability. Community health-promoting databases were used to identify potential subjects for each study setting. Potential participants in Study I were caregivers and children aged 6 months to 36 months who went to a health-promoting hospital for a vaccination or a health checkup. Potential participants for Study II were caregivers and children aged 6 months to 36 months, but they were contacted and provided with information regarding the study protocol via telephone. Participants for Study I were recruited from October 2018 to February 2019, with a total of 37 pairs agreeing to participate in the study. From December 2021 to February 2022, 34 pairs of participants were recruited for Study II.

Prior to data collection, participants were asked to provide written informed consent (Study I) or verbal consent (Study II). Throughout the study period, caregivers were required to have a smartphone connected to the internet and have or agree to apply for Facebook messenger. The child was required to have at least one tooth and have no serious medical problems affecting oral health status (eg, Down syndrome). Parent-child pairs were excluded from the study if they could not communicate in the Thai language or refused to consent to the study.

Interventions

After consenting to the study, the caregivers were trained to use the chatbots via their mobile devices. For Study I, caregivers were also given in-person instructions on toothbrushing technique, toothbrush selection, amount of toothpaste, and children's positioning while brushing teeth. The trainers used a doll to demonstrate toothbrushing (scrub technique); then, the caregiver practiced toothbrushing with their child. The training

session lasted 10 minutes to 15 minutes. As previously mentioned, due to the COVID-19 pandemic, in-person toothbrushing training was replaced with a video clip in Study II. The chatbot administrator observed the user engagement during the trial and assisted with resolving technical problems.

Data Collection and Outcome Assessment

There were 2 primary outcomes in this study: the effectiveness and usability of the chatbots. The chatbots' effectiveness was assessed in terms of changes in knowledge, oral health care perceptions based on PMT, and practices in oral hygiene care of young children. Both studies used a structured questionnaire designed to gather information on sociodemographic characteristics and oral health knowledge, perceptions, and practices at baseline and follow-up. Oral health knowledge questions comprised 11 items covering the following: appropriate time to begin brushing, frequency of brushing, brushing method, fluoride toothpaste, and child behavior management. A correct answer received 1 point, while an incorrect answer received 0 points. Questions about perceptions collected information about perceived severity, perceived vulnerability, response efficacy, and self-efficacy. The answer options were "positive perception" (3 points), "uncertain" (2 points), and "negative perception" (1 point). Oral hygiene practices were assessed using 4 categorical questions: brushing by a caregiver, frequency of toothbrushing per day, fluoride toothpaste usage, and the amount of toothpaste used.

The content validity and construct validity of the questionnaire were assessed by 2 experts for Study I and 3 experts for Study II. In both studies, face validity was determined through questionnaire-based pilot testing with 15 participants. A face-to-face interview by trained interviewers was used to collect data for Study I, while a self-administered online questionnaire via a Google Form was applied during the COVID-19 pandemic for Study II. In addition to the questionnaire, 1 dentist examined oral hygiene status under natural light using a disposable plastic straw and mouth mirror at baseline and follow-up in Study I. The amount of visible plaque found on the buccal surface of all erupted teeth was scored (0=no visible plaque; 1=presence of plaque).

Caregivers were asked to provide feedback on satisfaction with the chatbot at the last session (21st or 30th day). Open-ended questions were used to collect caregivers' suggestions. The overall satisfaction score rating ranged from 0 to 10 (0=very dissatisfied; 10=very satisfied). For Study II, an additional survey was administered using an in-depth interview (n=8). The usage patterns and days of engagement with the chatbots were recorded in log files (Figure 3).

Figure 3. Comparison of the methodologies used in the 2 study settings.

Items	Study I (Maikan)	Study II (Maelan)
Study design: pre-post design	✓	✓
Inclusion criteria	✓	✓
Exclusion criteria	✓	✓
Time between baseline and follow-up	1 month	2 months
Sample size	n=37	n=34
Sampling technique	✓	✓
Ethical consideration	Written informed consent	Verbal informed consent
Intervention	"21-Day FunDee" chatbot and hands-on toothbrushing practice with a doll	"30-Day FunDee" chatbot
Data collection	Face to face interview	Self-administered
Outcome assessments		
Effectiveness		
1. Change in oral health knowledge	✓	✓
2. Change in attitude	✓	✓
3. Change in practices	✓	✓
4. Change in visible plaque index	✓	✖
Usability of the chatbots	✓	✓
Overall satisfaction with the chatbots	✓	✓
Usage pattern	✓	✓

Data Analysis

The data were statistically analyzed using the free, open-access statistical software PSPP version 1.2.0 for Windows. Cronbach α was used to determine the questionnaire's internal consistency on overall oral health care perceptions based on PMT (α=.81 for Study I, α=.83 for Study II). Descriptive statistics were used to describe the baseline characteristics of study participants and responses to the questionnaires. The scores for perception and knowledge questions were normalized to 1 by summarizing the response scores of each question and dividing by the number of questions and the highest score. Then, means and standard deviation for each item and the overall score of the knowledge and perception questionnaires were calculated. The McNemar test was used to determine if there was a significant change in the proportions (before and after) of oral hygiene practices, and the Student paired *t* test was used to determine whether there was a statistically significant change in the mean scores for knowledge, perception, and plaque levels. The percentage of

caregivers who finished the chatbot session and the average number of days using chatbots were calculated.

Results

The sample comprised 71 pairs of participants: 37 for Study I and 34 for Study II. Similar sociodemographic characteristics were observed in the 2 studies (Table 1).

Both studies revealed a significant increase in toothbrushing for children by caregivers. Additionally, an increase in toothbrushing at least twice daily by caregivers for children was reported, but only Study I showed a significant difference (P=.02), with a 22% increase. The percentage of caregivers who used fluoride toothpaste and the amount of toothpaste used did not differ significantly. The plaque level significantly reduced after the intervention in Study I (Table 2).

Both the 21-day and 30-day chatbot applications significantly improved overall oral health care perceptions based on PMT, and every category, except perceived vulnerability, in Study II was not statistically different (Table 3).

Both chatbot applications statistically significantly increased overall knowledge and specific knowledge, such as brushing techniques and use of fluoride toothpaste. Additionally, the chatbot in Study I demonstrated a significantly greater understanding of the appropriate time to start toothbrushing and how to manage child behavior while brushing (Table 4).

Table 5 summarizes chatbot engagement and satisfaction with both chatbots. Satisfaction scores for Studies I and II were very high, at 9.2 and 8.6 out of 10, respectively. Almost all participants expressed that their respective chatbot was enjoyable, content-rich, capable of empathetic interaction, and worthy of assisting them improve their behaviors. Additionally, the users were impressed by the multimedia elements such as video and infographics for compelling storytelling and artwork. Quite a few participants raised concerns about the platform's stability. Some participants in Study I mentioned that in-person toothbrushing training made them more confident at toothbrushing and eager to clean their children's teeth.

Table 1. Characteristics of the samples in Studies I and II.

Characteristics	Study I (37 child-parent pairs)	Study II (34 child-parent pairs)
Child age (months), mean (SD)	16.4 (6.1)	20.1 (8.0)
Primary caregiver, n (%)		
Mother	35 (95)	30 (88)
Other	2 (6)	4 (12)
Caregiver age (years), mean (SD)	31.9 (9.0)	32.3 (10.1)
Caregiver education, n (%)		
Primary school	3 (8)	6 (18)
Junior high school	4 (11)	6 (18)
High school	16 (43)	10 (29)
Diploma or more	14 (38)	12 (35)
Caregiver occupation, n (%)		
Housewife	5 (13.5)	16 (47.1)
Employee	23 (62.2)	8 (23.5)
Agriculture	0 (0)	2 (5.9)
Business owner	3 (8.1)	3 (8.8)
Government	6 (16.2)	5 (14.7)
Religion, n (%)		
Buddhist	1 (3)	4 (12)
Muslim	36 (97)	30 (88)
Number of children in their house/siblings, mean (SD)	2.0 (1.1)	2.3 (1.5)

Table 2. The effects of the chatbot application on oral health care practices for young children in Studies I and II.

Oral health practices	Study I				Study II			
	Before, n (%)	After, n (%)	Chi-square *(df)*	*P* value	Before, n (%)	After, n (%)	Chi-square *(df)*	*P* value
Toothbrushing for children by caregiver								
Yes	28 (76)	36 (97)	6.40 (1)	.02[a]	22 (65)	30 (88)	5.33 (1)	.04[a]
No	9 (24)	1 (3)			12 (35)	4 (12)		
Frequency of brushing by caregiver								
Not everyday	13 (46)	3 (8)	12.23 (2)	.002	0 (0)	0 (0)	N/A[b]	.69[c]
Once a day	5 (18)	12 (33)			3 (14)	3 (10)		
Twice or more a day	10 (36)	21 (58)			19 (86)	27 (90)		
Fluoride toothpaste usage	22 (96)	36 (100)	N/A	.39[c]	13 (87)	25 (100)	N/A	.14[c]
Amount of toothpaste								
Smear	15 (65)	22 (61)	0.10 (1)	.75	16 (76)	25 (93)	N/A	.22[c]
Pea or regular size	8 (35)	14 (39)			5 (24)	2 (7)		
Plaque score[d]	0.48 (0.33)	0.18 (0.21)	6.82 (36)	<.001	N/A	N/A	N/A	N/A

[a]McNemar-test.

[b]N/A: not applicable.

[c]Fisher exact test.

[d]Mean (SD).

Table 3. The effects of the chatbot application on oral health care perceptions based on protection motivation theory in Studies I and II.

Perceptions[a]	Study I				Study II			
	Pretest, mean (SD)	Posttest, mean (SD)	*t* test *(df)*[b]	*P* value	Pretest, mean (SD)	Posttest, mean (SD)	*t* test *(df)*[b]	*P* value
Perceived severity	0.47 (0.33)	0.79 (0.26)	4.94 (36)	<.001	0.73 (0.20)	0.86 (0.15)	4.03 (33)	<.001
Perceived vulnerability	0.46 (0.51)	0.78 (0.42)	3.15 (36)	.003	0.83 (0.24)	0.89 (0.19)	1.07 (33)	.29
Response efficacy	0.57 (0.24)	0.90 (0.18)	7.76 (36)	<.001	0.68 (0.14)	0.79 (0.16)	3.89 (33)	<.001
Self-efficacy	0.71 (0.24)	0.89 (0.18)	4.15 (36)	<.001	0.72 (0.15)	0.80 (0.15)	2.64 (33)	.01
Overall perceptions	0.58 (0.19)	0.86 (0.16)	7.67 (36)	<.001	0.74 (0.12)	0.83 (0.12)	4.36 (33)	<.001

[a]The scores were normalized to 1.

[b]Paired *t* tests were used to compare the difference between pretest and posttest scores.

Table 4. The effects of the chatbot application on knowledge in Studies I and II.

Knowledge[a]	Study I				Study II			
	Pretest, mean (SD)	Posttest, mean (SD)	*t* test *(df)*[b]	*P* value	Pretest, mean (SD)	Posttest,mean (SD)	*t* test *(df)*[b]	*P* value
Appropriate time to start toothbrushing	0.22 (0.42)	0.86 (0.35)	7.33 (36)	.001	0.38 (0.49)	0.59 (0.50)	1.87 (33)	.07
Frequency of toothbrushing	0.78 (0.42)	0.80 (0.34)	1.00 (36)	.32	0.59 (0.31)	0.59 (0.34)	0.00 (33)	.99
Brushing method	0.80 (0.33)	1.00 (0)	3.83 (36)	.001	0.64 (0.17)	0.75 (0.17)	3.40 (33)	.002
Fluoride toothpaste	0.82 (0.27)	1.00 (0)	3.97 (36)	<.001	0.38 (0.49)	0.62 (0.49)	2.27 (33)	.03
Child behavior management	0.81 (0.29)	0.96 (0.14)	2.93 (36)	.006	0.69 (0.25)	0.72 (0.25)	1.00 (33)	.33
Overall knowledge	0.73 (0.21)	0.94 (0.09)	6.32 (36)	<.001	0.53 (0.26)	0.66 (0.23)	3.50 (33)	.001

[a]The scores were normalized to 1.

[b]Paired *t* tests were used to compare the difference between pretest and posttest scores.

Table 5. Engagement and satisfaction with the chatbot in Studies I and II.

Engagement and satisfaction	Study I	Study II
Chatbot engagement		
Full program engagement, n (%)	30 (81)	25 (74)
Days of engagement, mean (SD)	19.9 (4.9)	24.2 (2.8)
Days of engagement per week, mean (SD)	6.4 (1.5)	5.7 (1.7)
Satisfaction with the bot (0-10), mean (SD)	9.2 (0.9)	8.6 (1.2)

Discussion

To the authors' knowledge, this is the first study to use a chatbot-mediated intervention for oral health both prior to and during the COVID-10 pandemic. This study presents the effectiveness of both the 21-day FunDee chatbot with in-person training and the 30-day FunDee chatbot application in improving the child oral hygiene care by caregivers. Additionally, it demonstrated the potential for chatbots to serve as an alternative for in-person training for skills such as toothbrushing.

Both studies resulted in a significant improvement in toothbrushing by caregivers for children, with greater percentages reported than by other studies conducted in similar age groups utilizing traditional oral health education with or without in-person toothbrushing training [12]. Toothbrushing for children by caregivers has been shown as a key factor influencing the quality of plaque reduction for these young children compared with child self-toothbrushing [9].

In Study I, the percentage of caregivers who brushed their children's teeth twice daily increased from 36% to 58%, a significant increase of approximately 22%. In contrast, few changes were observed in Study II, possibly as a result of better baseline toothbrushing practices. In both studies, all caregivers eventually used fluoride toothpaste, which may result in long-term benefits for caries control [28]. Brushing teeth twice a day moderately increased, similar to a study with toddlers aged 9 months to 18 months in Thailand utilizing PMT with in-person toothbrushing training, which reported an increase from 11% to 42% over 1 year, compared with an increase from 11% to 16% in the control group receiving routine care [11]. This was consistent with a participatory approach intervention for caregivers of children younger than 6 years that included 90-minute, small-group sessions providing educational information, direct instruction, practice, and peer-to-peer problem solving. After 4 weeks to 8 weeks of intervention, improvement ranged from 59% to 89% [29]. It is worth noting that all other studies used human resources to accomplish the changes. In a meta-analysis, infrequent brushing had a significant impact on the incidence and increment of carious lesions in deciduous teeth (odds ratio 1.75, 95% CI 1.49-2.06) [30].

Study I demonstrated remarkable plaque reduction (62.5%) at the 1-month evaluation, which was similar to a study using a gamification application for mothers to reduce plaque accumulation in children aged 4 years to 5 years (plaque reduction of 50%) [31]. Both studies demonstrated a significant increase in oral health care knowledge, particularly regarding toothbrushing technique and the use of fluoride toothpaste. However, the caregivers believed that once per day was sufficient for brushing their children's teeth. It is necessary to gain a deeper understanding of caregivers' motivations and beliefs to improve chatbot conversations. Study I demonstrated slightly more progress in overall oral health care knowledge than Study II, which may be explained by the lower beginning knowledge score and more extended evaluation period in Study II.

PMT was used to develop the 2 chatbots aiming to improve toothbrushing behavior and engagement. Our 2 studies showed a high success rate in improving overall caregivers' perceptions except for perceived vulnerability, observed in Study I only. It is possible that, in Study I, in-person training raised the participants' awareness by showing them their children's level of plaque, while 30-Day FunDee persuaded caregivers to check their children's plaque and compare it with infographics provided by chatbot. To increase the motivation to change, Study II may have enhanced perceived vulnerability using AI technology to compare plaque levels based on photographs of each participant that were more contextually relevant.

In-person training is a powerful way to increase one's ability and empower confidence in toothbrushing [13,14]. A study demonstrated that increased perceived severity and self-efficacy via in-person training for toothbrushing techniques can reduce plaque levels on a long-term basis [11]. Additionally, Finlayson et al [32] found that increased maternal oral health–related self-efficacy is associated with increased frequency of toothbrushing in children aged 1 year to 5 years. Self-efficacy is a crucial aspect of PMT and has been identified as a factor that enables individuals to adhere to healthier behaviors and predicts a range of health behaviors including oral self-care [33-35]. In our study, we discovered that the chatbot, 30-Day FunDee, effectively improves toothbrushing practices, self-efficacy, response efficacy, and threat perceptions in terms of perceived severity. Thus, this chatbot might be used as an alternative for in-person training for toothbrushing for young children by caregivers.

Interestingly, both studies had high engagement rates (6.4 days and 5.7 days per week in Study I and Study II, respectively) compared with other studies on chatbots in health, such as the study by Jang et al [36] that reported 5.1 days per week of chatbot engagement with a 4-week chatbot course and the study by Fitzpatrick et al [37] that reported 6.1 days per week of chatbot adherence with a 14-day chatbot course. It is noticeable that a longer period for chatbot delivery is related to a decreased engagement rate. Engagement with the chatbot in our study

may have resulted from the high satisfaction score for the chatbots. The chatbots were satisfying since they contain attractive multimedia; understandable content; friendly, empathetic dialogue; and utility as well as being easy to use. This is consistent with other studies showing high acceptability via chat enjoyment, bonding, creation of social and emotional relationships, ease of use, usefulness, and a desire to use [38,39].

Based on our experience, the model by Zhang et al [20] is useful for planning the overall conversational flow and creating more humanized chatbots. Incorporating content and behavioral change theory into the conversational flow was the most challenging aspect of achieving harmony. The conversational flow reflected, to some extent, what we had used to successfully motivate patients to improve their oral health behavior and what we had learned by applying theory to practice.

This secondary data analysis has some limitations. First, both studies used a pre-post design that may have a maturity bias; therefore, the chatbot's effectiveness in improving oral health behavior may be overestimated. Second, although our study was conducted with similar research methodology, the interview procedure and follow-up period differed. Study I used face-to-face interviews and a shorter follow-up period of 9 days after the intervention ended; therefore, the study may be subject to examiner bias, and the results may reflect the chatbot's short-term effect. Although Study II used a self-administered online questionnaire, its validity could be compromised if individuals responded unintentionally, and a longer follow-up time could influence memory retention. To generalize the results of this study, randomized trials in different groups and a longer-term evaluation of caries prevention should be conducted in future studies. Furthermore, to improve the effectiveness of chatbots, adaptive learning and AI-based conversation should be incorporated.

This study introduced chatbot applications as a new normal approach to oral health education. We demonstrated the effectiveness of using chatbots to empower caregivers of young children to perform oral hygiene care for the child prior to and during the COVID-19 pandemic and showed the possibility of using chatbots to improve toothbrushing abilities without actual in-person training.

Acknowledgments

This research was supported by the Faculty of Dentistry fund. The authors are grateful for the tremendous support from Dr Narim Tokanee, Dr Rohannee Yalaeka, and all the enthusiastic staff from Maikan and Maelan hospitals. We appreciate Nawaporn Lukkanatinnaporn for her incredible work on the graphic design and dialogue in creating the chatbot 21-Day FunDee. Finally, many thanks to the caregivers who actively participated in our research.

Authors' Contributions

JH and SP performed the majority of the work on the 21-Day FunDee chatbot including studying the literature, creating the chatbot, and gathering and analyzing data. KP, JH, and SP created the 30-Day FunDee chatbot by conducting a literature review, developed the chatbot, and collected and analyzed data. SN contributed to the research design for Study II. SN, JH, SP, and KP conceptualized and designed the presentation, performed the data comparison analysis, and wrote the first draft of the paper. JH and SN edited the paper to produce the final manuscript.

Conflicts of Interest

None declared.

References

1. Zaror C, Matamala-Santander A, Ferrer M, Rivera-Mendoza F, Espinoza-Espinoza G, Martínez-Zapata MJ. Impact of early childhood caries on oral health-related quality of life: A systematic review and meta-analysis. Int J Dent Hyg 2022 Feb;20(1):120-135. [doi: 10.1111/idh.12494] [Medline: 33825317]
2. Filstrup SL, Briskie D, da Fonseca M, Lawrence L, Wandera A, Inglehart MR. Early childhood caries and quality of life: child and parent perspectives. Pediatr Dent 2003;25(5):431-440. [Medline: 14649606]
3. Thomas CW, Primosch RE. Changes in incremental weight and well-being of children with rampant caries following complete dental rehabilitation. Pediatr Dent 2002;24(2):109-113. [Medline: 11991312]
4. Casamassimo PS, Thikkurissy S, Edelstein BL, Maiorini E. Beyond the dmft: the human and economic cost of early childhood caries. J Am Dent Assoc 2009 Jun;140(6):650-657. [doi: 10.14219/jada.archive.2009.0250] [Medline: 19491160]
5. Chaffee B, Featherstone J, Gansky S, Cheng J, Zhan L. Caries risk assessment item importance: risk designation and caries status in children under age 6. JDR Clin Trans Res 2016 Jul 05;1(2):131-142 [FREE Full text] [doi: 10.1177/2380084416648932] [Medline: 27403458]
6. Yoon RK, Smaldone AM, Edelstein BL. Early childhood caries screening tools: a comparison of four approaches. J Am Dent Assoc 2012 Jul;143(7):756-763 [FREE Full text] [doi: 10.14219/jada.archive.2012.0263] [Medline: 22751977]
7. Poureslami HR, Van Amerongen WE. Early childhood caries (ECC): an infectious transmissible oral disease. Indian J Pediatr 2009 Feb 4;76(2):191-194. [doi: 10.1007/s12098-008-0216-1] [Medline: 19082537]

8. Chanpum P, Duangthip D, Trairatvorakul C, Songsiripradubboon S. Early childhood caries and its associated factors among 9- to 18-month old exclusively breastfed children in Thailand: a cross-sectional study. Int J Environ Res Public Health 2020 May 04;17(9):1 [FREE Full text] [doi: 10.3390/ijerph17093194] [Medline: 32375351]
9. Phantumvanit P, Makino Y, Ogawa H, Rugg-Gunn A, Moynihan P, Petersen PE, et al. WHO global consultation on public health intervention against early childhood caries. Community Dent Oral Epidemiol 2018 Jun;46(3):280-287. [doi: 10.1111/cdoe.12362] [Medline: 29380407]
10. Dental HB. The 8th national oral health survey report. Thailand. 2018. URL: http://www.dent.chula.ac.th/upload/news/791/file_1_5834.pdf [accessed 2022-07-01]
11. Kimhasawad W, Punyanirun K, Somkotra T, Detsomboonrat P, Trairatvorakul C, Songsiripradubboon S. Comparing protection-motivation theory-based intervention with routine public dental health care. Int J Dent Hyg 2021 Aug;19(3):279-286. [doi: 10.1111/idh.12522] [Medline: 34014618]
12. Jiang EM, Lo ECM, Chu CH, Wong MCM. Prevention of early childhood caries (ECC) through parental toothbrushing training and fluoride varnish application: a 24-month randomized controlled trial. J Dent 2014 Dec;42(12):1543-1550. [doi: 10.1016/j.jdent.2014.10.002] [Medline: 25448437]
13. Ungchusak C. Oral health promotion and prevention of early childhood caries. Thai Dental Public Health Journal 2017;22:44-61 [FREE Full text]
14. Thanakanjanaphakdee W, Trairatvorakul C. Effectiveness of parental hands-on toothbrushing instruction toward the 1-year incremental dmf rate of 9-18 month old children. J Dent Assoc Thai 2010;60(2):85-94 [FREE Full text]
15. 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]
16. Parmar P, Ryu J, Pandya S, Sedoc J, Agarwal S. Health-focused conversational agents in person-centered care: a review of apps. NPJ Digit Med 2022 Feb 17;5(1):21 [FREE Full text] [doi: 10.1038/s41746-022-00560-6] [Medline: 35177772]
17. Gabarron E, Larbi D, Denecke K, Årsand E. What do we know about the use of chatbots for public health? Stud Health Technol Inform 2020 Jun 16;270:796-800. [doi: 10.3233/SHTI200270] [Medline: 32570492]
18. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
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. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
21. Rogers RW. A protection motivation theory of fear appeals and attitude change. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
22. Maddux JE, Rogers RW. Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. Journal of Experimental Social Psychology 1983 Sep;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9]
23. Milne S, Sheeran P, Orbell S. Prediction and intervention in health-related behavior: a meta-analytic review of protection motivation theory. J Appl Social Pyschol 2000 Jan;30(1):106-143 [FREE Full text] [doi: 10.1111/j.1559-1816.2000.tb02308.x]
24. Kim J, Yang K, Min J, White B. Hope, fear, and consumer behavioral change amid COVID-19: Application of protection motivation theory. Int J Consum Stud 2022 Mar;46(2):558-574 [FREE Full text] [doi: 10.1111/ijcs.12700] [Medline: 34220343]
25. Lahiri A, Jha SS, Chakraborty A, Dobe M, Dey A. Role of threat and coping appraisal in protection motivation for adoption of preventive behavior during COVID-19 pandemic. Front Public Health 2021;9:678566 [FREE Full text] [doi: 10.3389/fpubh.2021.678566] [Medline: 34291027]
26. Floyd DL, Prentice-dunn S, Rogers RW. A meta-analysis of research on protection motivation theory. J Appl Social Pyschol 2000 Feb;30(2):407-429 [FREE Full text] [doi: 10.1111/j.1559-1816.2000.tb02323.x]
27. Toniazzo MP, Nodari D, Muniz FWMG, Weidlich P. Effect of mHealth in improving oral hygiene: A systematic review with meta-analysis. J Clin Periodontol 2019 Mar 19;46(3):297-309. [doi: 10.1111/jcpe.13083] [Medline: 30761580]
28. Boustedt K, Dahlgren J, Twetman S, Roswall J. Tooth brushing habits and prevalence of early childhood caries: a prospective cohort study. Eur Arch Paediatr Dent 2020 Feb;21(1):155-159. [doi: 10.1007/s40368-019-00463-3] [Medline: 31338770]
29. Huebner CE, Milgrom P. Evaluation of a parent-designed programme to support tooth brushing of infants and young children. Int J Dent Hyg 2015 Feb;13(1):65-73 [FREE Full text] [doi: 10.1111/idh.12100] [Medline: 25070036]
30. Kumar S, Tadakamadla J, Johnson NW. Effect of toothbrushing frequency on incidence and increment of dental caries: a systematic review and meta-analysis. J Dent Res 2016 Oct;95(11):1230-1236. [doi: 10.1177/0022034516655315] [Medline: 27334438]

31. Zolfaghari M, Shirmohammadi M, Shahhosseini H, Mokhtaran M, Mohebbi SZ. Development and evaluation of a gamified smart phone mobile health application for oral health promotion in early childhood: a randomized controlled trial. BMC Oral Health 2021 Jan 07;21(1):18 [FREE Full text] [doi: 10.1186/s12903-020-01374-2] [Medline: 33413304]
32. Finlayson TL, Siefert K, Ismail AI, Sohn W. Maternal self-efficacy and 1-5-year-old children's brushing habits. Community Dent Oral Epidemiol 2007 Aug;35(4):272-281. [doi: 10.1111/j.1600-0528.2007.00313.x] [Medline: 17615014]
33. Buglar ME, White KM, Robinson NG. The role of self-efficacy in dental patients' brushing and flossing: testing an extended Health Belief Model. Patient Educ Couns 2010 Feb;78(2):269-272. [doi: 10.1016/j.pec.2009.06.014] [Medline: 19640670]
34. Mizutani S, Ekuni D, Furuta M, Tomofuji T, Irie K, Azuma T, et al. Effects of self-efficacy on oral health behaviours and gingival health in university students aged 18- or 19-years-old. J Clin Periodontol 2012 Sep;39(9):844-849. [doi: 10.1111/j.1600-051X.2012.01919.x] [Medline: 22780323]
35. Stewart JE, Wolfe GR, Maeder L, Hartz GW. Changes in dental knowledge and self-efficacy scores following interventions to change oral hygiene behavior. Patient Educ Couns 1996 Apr;27(3):269-277. [doi: 10.1016/0738-3991(95)00843-8] [Medline: 8788355]
36. 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]
37. 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]
38. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al. Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. In: Maedche A, vom Brocke J, Hevner A, editors. Designing the Digital Transformation. Cham, Switzerland: Springer International Publishing; 2017.
39. 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]

Abbreviations

AI: artificial intelligence
ECC: early childhood caries
PMT: protection motivation theory

Edited by T Leung, V Arnold, T Purnat; submitted 16.06.22; peer-reviewed by J Zhang, S Sarejloo, P Phantumvanit; comments to author 17.08.22; revised version received 02.09.22; accepted 27.09.22; published 21.10.22.

Please cite as:
Pithpornchaiyakul S, Naorungroj S, Pupong K, Hunsrisakhun J
Using a Chatbot as an Alternative Approach for In-Person Toothbrushing Training During the COVID-19 Pandemic: Comparative Study
J Med Internet Res 2022;24(10):e39218
URL: https://www.jmir.org/2022/10/e39218
doi:10.2196/39218
PMID:36179147

Research Letter

An Increase in Antibiotic Prescribing for Respiratory Tract Infections Through Telehealth Consultations: Retrospective Study in Australian General Practice

Chisato Imai[1*], BSc, MPH, PhD; Janaki Amin[2*], BSc, MPH, PhD; Mirela Prgomet[1*], PhD; Christopher Pearce[1,3*], MBBS, MFA, PhD; Andrew Georgiou[1*], BSc, PhD

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

[2]Department of Health Sciences, Macquarie University, Sydney, Australia

[3]Aurora Primary Care Research Institute, Blackburn, Australia

[*]all authors contributed equally

Corresponding Author:
Chisato Imai, BSc, MPH, PhD
Australian Institute of Health Innovation
Macquarie University
75 Talavera Road
Level 6
Sydney, 2109
Australia
Phone: 61 02 9850 2415
Email: chrissy.imai@mq.edu.au

(J Med Internet Res 2022;24(10):e40876) doi:10.2196/40876

KEYWORDS

general practice; anti-infective agents; antibiotics; medication; prescriptions; respiratory tract infections; infection; telehealth; telemedicine; Australia

Introduction

General practitioners (GPs) are the first point of contact for most people in Australia seeking medical attention. They play an important role in the prevention, early detection, and management of both acute and chronic diseases. In Australian general practice, antibiotics are commonly prescribed for respiratory tract infections (RTIs) despite evidence of limited efficacy [1,2]. A previous study identified that the antibiotic prescribing rates for RTIs by GPs are up to 9 times higher than recommended by practice guidelines [3]. Such findings have raised public health concerns about overprescribing antibiotics, emphasizing the need for continued monitoring of antibiotic prescribing activities in Australian general practice.

In mid-March 2020, the Australian government implemented the expansion of telehealth services covered by Medicare (Australia's universal health insurance) in response to the COVID-19 pandemic. Subsequently, many GPs shifted care delivery from face-to-face consultations to telehealth (telephone or video-conference consultations). A study conducted in the early stage of the pandemic showed lower rates of medication prescriptions via telehealth compared to face-to-face consultations in Australian general practice [4]; however, antibiotic prescribing for RTIs by consultation modality is yet to be explored. Therefore, we examined antibiotic prescribing for RTIs via telehealth in comparison with face-to-face consultations from April 2020 to November 2021.

Methods

Data and Analysis

This retrospective study used data from the Population Level Analysis and Reporting (POLAR) platform [4]. POLAR comprises deidentified electronic health records collated from approximately 1000 general practices across 2 states in Australia (ie, Victoria and New South Wales). To obtain our variables of interest, diagnosis data were used to identify consultations for respiratory infections based on SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms) codes (respiratory infection, [viral/bacterial/recurrent/acute] upper respiratory tract infection, and [viral/bacterial/recurrent/acute] lower respiratory tract infection); antibiotic prescriptions were identified via the Anatomical Therapeutic Chemical code J01 (antibacterials for systemic use) in prescription data; and consultation modality was defined based on service item numbers from Medicare billing data [5].

For the analysis, we examined the mean weekly percentage of consultations with antibiotic prescribing between April 1, 2020, and November 30, 2021. We also determined the patient-level probability of an antibiotic being prescribed during consultations in the periods June to November 2020 and June to November 2021, using generalized estimating equations with the Huber-White standard error, adjusted for patient factors (age, gender, remoteness, state of residence, active status as defined by the Royal Australian College of General Practitioners [RACGP], and recent history of consultation) and a sampling effect within the general practices. Active patients (ie, regular visitors to a practice) were defined based on the definition from the RACGP as individuals who had attended the practices 3 or more times in the past 2 years at the time of visit [6].

Ethics Approval

Ethics approval was obtained from the Macquarie University Human Research Ethics Committee (#52020675617176).

Results

Between April 2020 and November 2021, a total of 105,719 individuals had an RTI diagnosis and attended 141,444 consultations: 92,318 (65.3%) face-to-face visits and 49,126 (34.7%) telehealth consultations (comprising 48,159 via telephone and 967 via video conference).

In 2020, the weekly mean antibiotic prescribing rates for RTIs were 56.7% (95% CI 54.2%-59.3%) for face-to-face consultations and 40.8% (95% CI 37.2%-44.4%) for telehealth (Figure 1). In 2021, the weekly mean prescribing rates were 58.6% (95% CI 55.8%-61.4%) for face to face and 61.0% (95% CI 59.1%-63.0%) for telehealth. We also evaluated the number of prescriptions during the study period and observed the same longitudinal trend.

At the patient level, the probability of receiving an antibiotic prescription through a telehealth consultation also increased from 59.3% (95% CI 57.6%-61.0%) in 2020 to 65.7% (95% CI 64.4%-67.0%) in 2021 (Table 1). The probability via face-to-face consultations was consistent across 2020 and 2021: 65.2% (95% CI 63.4%-66.9%) and 66.9% (95% CI 65.5%-68.2%), respectively.

Figure 1. Weekly mean percentage of respiratory tract infection consultations with an antibiotic prescription for F2F and telehealth consultations (April 2020 to November 2021). F2F: face-to-face; NSW: New South Wales; VIC: Victoria.

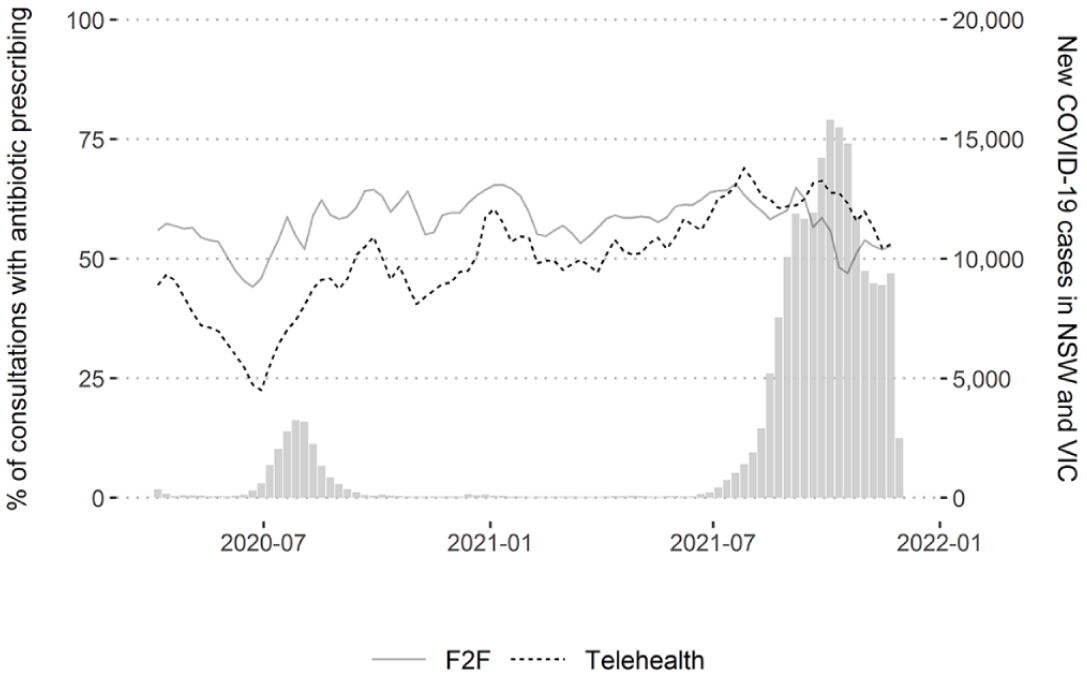


Table 1. Patient probability of receiving antibiotic prescriptions in a consultation for respiratory tract infection.

Variables	Estimated probability (%) of receiving antibiotic prescription (95% CI)	
	2020	2021
Age group (years)		
<10	54.2 (52.0-56.6)	54.7 (53.0-56.4)
10-24	50.5 (47.6-53.6)	59.1 (56.3-62.0)
25-44	61.6 (58.9-64.5)	67.7 (65.9-69.6)
45-64	74.6 (71.2-78.1)	73.2 (71.0-75.5)
≥65	93.4 (87.7-99.4)	83.3 (80.3-86.5)
Sex		
Female	65.2 (63.4-66.9)	66.9 (65.5-68.2)
Male	58.0 (56.2-59.8)	65.6 (64.1-67.1)
Active status		
Nonactive	48.9 (47.2-50.7)	51.9 (50.3-53.6)
Active	65.2 (63.4-66.9)	66.9 (65.5-68.2)
Remoteness		
Major cities	65.2 (63.4-66.9)	66.9 (65.5-68.2)
Regional or remote areas	75.8 (72.1-79.7)	77.3 (74.9-79.7)
Previous encounter (≤2 weeks)		
No	65.2 (63.4-66.9)	66.9 (65.5-68.2)
Yes	63.6 (61.9-65.3)	66.5 (65.2-67.9)
State		
New South Wales	65.2 (63.4-66.9)	67.7 (66.4-69.1)
Victoria	54.6 (52.9-56.4)	66.9 (65.5-68.2)
Consultation		
Face to face	65.2 (63.4-66.9)	66.9 (65.5-68.2)
Telehealth	59.3 (57.6-61.0)	65.7 (64.4-67.0)

Discussion

Antibiotic prescribing via telehealth increased over time, with rates initially much lower than face-to-face consultations; however, the prescribing rates between the two consultation modalities became equivalent toward the end of 2021.

While telehealth offers some advantages such as less travel time and prevention of infectious disease transmissions, the caveats include limited physical examination capabilities. Such limitations may potentially impact the adequacy of medication prescribing, particularly in populations like children who have difficulty verbalizing symptoms.

Considering that high antibiotic prescribing rates for RTIs by Australian GPs is a long-standing public health concern, the findings from this study highlight the need for monitoring the impacts of telehealth on medication prescribing in general practice. Further, studies on telehealth decision-making processes for antibiotic prescribing that evaluate prescribing adequacy appear critical.

Acknowledgments

We would like to thank and acknowledge all general practitioners, participating Primary Health Networks, as well as the data custodian Outcome Health. This research was also supported by Digital Health CRC Limited, which is funded under the Commonwealth's Cooperative Research Centres Program.

Conflicts of Interest

None declared.

References

1. Del Mar CB, Scott AM, Glasziou PP, Hoffmann T, van Driel ML, Beller E, et al. Reducing antibiotic prescribing in Australian general practice: time for a national strategy. Med J Aust 2017 Nov 6;207(9):401-406. [doi: 10.5694/mja17.00574] [Medline: 29092694]
2. AURA 2016: first Australian report on antimicrobial use and resistance in human health. Australian Commission on Safety and Quality in Health Care. 2016. URL: https://www.safetyandquality.gov.au/publications-and-resources/resource-library/aura-2016-first-australian-report-antimicrobial-use-and-resistance-human-health-full-report [accessed 2022-03-30]
3. McCullough AR, Pollack AJ, Plejdrup Hansen M, Glasziou PP, Looke DF, Britt HC, et al. Antibiotics for acute respiratory infections in general practice: comparison of prescribing rates with guideline recommendations. Med J Aust 2017 Jul 17;207(2):65-69. [doi: 10.5694/mja16.01042] [Medline: 28701117]
4. Wabe N, Thomas J, Sezgin G, Sheikh MK, Gault E, Georgiou A. Medication prescribing in face-to-face versus telehealth consultations during the COVID-19 pandemic in Australian general practice: a retrospective observational study. BJGP Open 2022 Mar;6(1):6 [FREE Full text] [doi: 10.3399/BJGPO.2021.0132] [Medline: 34819296]
5. COVID-19 telehealth MBS items. Department of Health and Aged Care, Australian Government. 2020 Aug 13. URL: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/news-2020-03-29-latest-news-March [accessed 2022-09-07]
6. Standards for general practices, 4th edition. The Royal Australian College of General Practitioners. URL: https://www.racgp.org.au/running-a-practice/practice-standards/standards-4th-edition [accessed 2022-03-30]

Abbreviations

GP: general practitioner
POLAR: Population Level Analysis and Reporting
RTI: respiratory tract infection
RACGP: Royal Australian College of General Practitioners
SNOMED-CT: Systematized Nomenclature of Medicine – Clinical Terms

Edited by T Leung; submitted 08.07.22; peer-reviewed by G Tramper-Stranders, S Leitch; comments to author 31.08.22; revised version received 14.09.22; accepted 21.09.22; published 18.10.22.

Please cite as:
Imai C, Amin J, Prgomet M, Pearce C, Georgiou A
An Increase in Antibiotic Prescribing for Respiratory Tract Infections Through Telehealth Consultations: Retrospective Study in Australian General Practice
J Med Internet Res 2022;24(10):e40876
URL: https://www.jmir.org/2022/10/e40876
doi:10.2196/40876
PMID:36256826

Research Letter

The Impact of Social Isolation, Loneliness, and Technology Use During the COVID-19 Pandemic on Health-Related Quality of Life: Observational Cross-sectional Study

Eric Balki[1], BSc, MSc, MA, PhD; Niall Hayes[2], BSc, MSc, PhD; Carol Holland[1], BSc, PhD

[1]Division of Health Research, Lancaster University, Lancaster, United Kingdom

[2]Directorate, Nottingham Trent University, Nottingham, United Kingdom

Corresponding Author:
Eric Balki, BSc, MSc, MA, PhD
Division of Health Research
Lancaster University
Furness Building
Hazelrigg Ln
Lancaster, LA1 4YG
United Kingdom
Phone: 44 7801972693
Email: e.balkhi@lancaster.ac.uk

(J Med Internet Res 2022;24(10):e41536) doi:10.2196/41536

KEYWORDS

health-related quality of life; healthy aging; older adult; elder; older person; older population; geriatric; gerontology; technology intervention; COVID-19; pandemic; loneliness; social isolation; isolation; isolated; lonely; cross-sectional; technology use; digital literacy; acceptance

Introduction

Health-related quality of life (HRQoL), defined as a person's self-perceived health status in relation to their social, cultural, and environmental context, is linked to better health and the ability to deal with adverse life events [1]. Social factors such as loneliness are known to influence HRQoL negatively [2]. The COVID-19 pandemic has disproportionately impacted older adults, with social distancing measures worsening isolation levels [3], which we hypothesize has resulted in lower levels of HRQoL (hypothesis 1).

Further, technology use is linked to improved self-rated health and psychological well-being, alleviating loneliness among older adults, and encouraging behaviors that may lead to better levels of HRQoL [4]. Digital communication tools became critical during the pandemic to remain socially connected and helped prevent social health risks [5], potentially benefiting those with lower HRQoL [6]. We hypothesized that technology use could predict higher HRQoL (hypothesis 2). Moreover, disease containment measures resulted in increased isolation and loneliness among older adults [3], which could impact HRQoL (hypothesis 3). Increased knowledge about how HRQoL was impacted by pandemic loneliness, isolation, and technology use may better inform health care workers, policy makers, and the public.

Methods

This was an observational cross-sectional study from March 16, 2020, to June 21, 2021, when social distancing mandates were in force. Participants were recruited in England.

Ethics Approval

The study received ethical approval from the University Research Ethics Committee (Ref FHMREC19121).

Participants

Eligible participants were living in their own homes, proficient in English, and aged ≥65 years. The sample (G*Power confirmed effect size of 87) consisted of 89 people aged 65 to 92 (mean 73.2, SD 7.46) years.

Variables and Measures

Participants completed a background questionnaire capturing age, gender, and ethnicity. We used the following standardized measures: UCLA Loneliness Scale [7], Technology Experience Questionnaire [8], Lubben's Social Isolation Scale, and Short-Form 36 [9], a measure of HRQoL comprising eight health scales (physical/mental).

Procedure

Surveys were conducted via telephone, with further analysis done using SPSS Ver 28 (IBM Corp).

Statistical Methods

Higher scores on the UCLA Loneliness Scale and technology use measures indicated greater loneliness and technology use; lower scores on Lubben's scale indicated greater isolation. Pearson correlation determined whether lower social isolation (hypothesis 1) and greater technology use (hypothesis 2) were associated with higher HRQoL. Multiple linear regression models were built to evaluate whether loneliness predicted HRQoL after controlling for social isolation and technology use (hypothesis 3).

Results

Low social isolation (hypothesis 1) and higher technology use (hypothesis 2) were significantly associated with higher HRQoL (Table 1).

Multiple linear regression was calculated (Table 2) for hypothesis 3. Model 1, incorporating loneliness, explained 24.9% of the variance in HRQoL. Model 2, incorporating social isolation, explained an additional nonsignificant 0.1% of the variance ($F_{1,89}$=0.112; P=.74). Model 3, adding technology use, explained an additional 5.5% of the variance ($F_{1,88}$=6.93; P=.01).

Semipartial correlations squared showed unique amount of variance; only technology use predicted a significant unique amount of the variance in HRQoL (sr^2=0.0547; P=.01), followed by loneliness (sr^2=0.0179; P=.14) and social isolation (sr^2=0.0004; P=.82).

Table 1. Correlational analysis between variables (N=89).

	UCLA Loneliness score	HRQoL[a]	Technology use	Social isolation
UCLA Loneliness score				
Pearson correlation	—[b]	—0.499	–0.631	–0.853
P value	—	<.001	<.001	<.001
HRQoL				
Pearson correlation	−0.499	—	0.497	0.442
P value	<.001	—	<.001	<.001
Technology use				
Pearson correlation	−0.631	0.497	—	0.577
P value	<.001	<.001	—	<.001
Social isolation				
Pearson correlation	−0.853	0.442	0.557	—
P value	<.001	<.001	<.001	—

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

[b]Not applicable.

Table 2. Model output and coefficients of multiple linear regression models for health-related quality of life (N=89).

Independent variables	Model 1			Model 2			Model 3		
	b (SE)	B	*P* value	b (SE)	B	*P* value	b (SE)	B	*P* value
Loneliness	−4.07 (0.745)	−0.499	<.001	−3.66 (1.436)	−0.449	.01	−2.246 (1.490)	−0.275	.14
Social isolation	N/A[a]	N/A	N/A	0.559 (1.671)	0.059	.74	0.369 (1.62)	0.039	.82
Technology use	N/A	N/A	N/A	N/A	N/A	N/A	1.071 (0.407)	0.302	.01
Intercept	757.851 (37.75)	N/A	<.001	723.318 (109.926)	N/A	<.001	536.117 (128.009)	N/A	<.001
R^2 (ΔR^2)	0.249	N/A	<.001	0.250 (0.001)	N/A	.74	0.305 (0.055)	N/A	.01
F test (*df*)	29.871 (1,90)	N/A	<.001	14.844 (2,89)	N/A	<.001	12.865 (3,88)	N/A	<.001

[a]N/A: not applicable.

Discussion

Few studies to date have examined the impact of social isolation, loneliness, and technology use together on HRQoL in older adults in England during the pandemic. We found that loneliness negatively impacts HRQoL, and technology use positively impacts it. Although social isolation has been linked to HRQoL, it had a low impact when loneliness was accounted for.

Technology use was related to higher HRQoL, aligning our findings with the results of previous studies [9]. However, the magnitude of the positive effect was notable when considering prepandemic studies [10]. Loneliness impacted HRQoL even when social isolation and technology use were accounted for, in agreement with previous literature [10]. The cross-sectional design prevented us from determining causality and was the main limitation of this study. Our study has relevant implications for health professionals such as health psychologists seeking to improve the HRQoL of older adults, especially through adverse life events like the pandemic or other circumstances that would put older adults in a similar situation where their mobility has been restricted. Our study informs that loneliness should be addressed, in conjunction with increasing technology use, in interventions. The absence of longitudinal studies examining the same cohort before and after the pandemic makes this interpretation speculative. Further research is needed to determine causes, and future studies need to examine pandemic-linked long-term impacts on the mental health and well-being of older adults.

Conflicts of Interest

None declared.

References

1. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7] [Medline: 9817135]
2. Salihu HM, Adegoke K, Turner D, Al Agili D, Berry EL. Social support and health-related quality of life among low-income women: findings from community-based participatory research. South Med J 2017 Apr;110(4):270-277. [doi: 10.14423/SMJ.0000000000000635] [Medline: 28376524]
3. Groarke JM, Berry E, Graham-Wisener L, McKenna-Plumley PE, McGlinchey E, Armour C. Loneliness in the UK during the COVID-19 pandemic: cross-sectional results from the COVID-19 Psychological Wellbeing Study. PLoS One 2020;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
4. Goethals L, Barth N, Guyot J, Hupin D, Celarier T, Bongue B. Impact of home quarantine on physical activity among older adults living at home during the COVID-19 pandemic: qualitative interview study. JMIR Aging 2020 May 07;3(1):e19007 [FREE Full text] [doi: 10.2196/19007] [Medline: 32356777]
5. Balki E, Hayes N, Holland C. Effectiveness of technology interventions in addressing social isolation, connectedness, and loneliness in older adults: a systematic umbrella review. JMIR Aging (forthcoming) 2022:1.
6. Ang S, Chen TY. Going online to stay connected: online social participation buffers the relationship between pain and depression. J Gerontol B Psychol Sci Soc Sci 2019 Aug 21;74(6):1020-1031. [doi: 10.1093/geronb/gby109] [Medline: 30260444]
7. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
8. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
9. Khalaila R, Vitman-Schorr A. Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: mediating and moderating effects. Qual Life Res 2018 Feb;27(2):479-489. [doi: 10.1007/s11136-017-1749-4] [Medline: 29210015]
10. Berg-Weger M, Morley JE. Editorial: loneliness and social isolation in older adults during the COVID-19 pandemic: implications for gerontological social work. J Nutr Health Aging 2020;24(5):456-458 [FREE Full text] [doi: 10.1007/s12603-020-1366-8] [Medline: 32346678]

Abbreviations

HRQoL: health-related quality of life.

Edited by G Eysenbach, T Leung; submitted 29.07.22; peer-reviewed by B Wang, H Lum, SGS Shah; comments to author 23.08.22; revised version received 15.09.22; accepted 08.10.22; published 19.10.22.

Please cite as:
Balki E, Hayes N, Holland C
The Impact of Social Isolation, Loneliness, and Technology Use During the COVID-19 Pandemic on Health-Related Quality of Life: Observational Cross-sectional Study
J Med Internet Res 2022;24(10):e41536
URL: https://www.jmir.org/2022/10/e41536
doi:10.2196/41536
PMID:36260401

Original Paper

Tracking the Impact of COVID-19 and Lockdown Policies on Public Mental Health Using Social Media: Infoveillance Study

Minghui Li[1,2*], BA; Yining Hua[3,4*], BA; Yanhui Liao[5], MD; Li Zhou[3,4], MD, PhD; Xue Li[1,2], PhD; Ling Wang[6], PhD; Jie Yang[1,2], PhD

[1]Department of Big Data in Health Science School of Public Health, Center of Clinical Big Data and Analytics of The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]The Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province, Hangzhou, China

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[4]Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

[5]Department of Psychiatry, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[6]Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Jie Yang, PhD
Department of Big Data in Health Science School of Public Health, Center of Clinical Big Data and Analytics of The Second Affiliated Hospital
Zhejiang University School of Medicine
866 Yuhangtang Road
Hangzhou, 310058
China
Phone: 86 19157731185
Fax: 86 571 87077982
Email: jieynlp@gmail.com

Abstract

Background: The COVID-19 pandemic and its corresponding preventive and control measures have increased the mental burden on the public. Understanding and tracking changes in public mental status can facilitate optimizing public mental health intervention and control strategies.

Objective: This study aimed to build a social media–based pipeline that tracks public mental changes and use it to understand public mental health status regarding the pandemic.

Methods: This study used COVID-19–related tweets posted from February 2020 to April 2022. The tweets were downloaded using unique identifiers through the Twitter application programming interface. We created a lexicon of 4 mental health problems (depression, anxiety, insomnia, and addiction) to identify mental health–related tweets and developed a dictionary for identifying health care workers. We analyzed temporal and geographic distributions of public mental health status during the pandemic and further compared distributions among health care workers versus the general public, supplemented by topic modeling on their underlying foci. Finally, we used interrupted time series analysis to examine the statewide impact of a lockdown policy on public mental health in 12 states.

Results: We extracted 4,213,005 tweets related to mental health and COVID-19 from 2,316,817 users. Of these tweets, 2,161,357 (51.3%) were related to "depression," whereas 1,923,635 (45.66%), 225,205 (5.35%), and 150,006 (3.56%) were related to "anxiety," "insomnia," and "addiction," respectively. Compared to the general public, health care workers had higher risks of all 4 types of problems (all $P<.001$), and they were more concerned about clinical topics than everyday issues (eg, "students' pressure," "panic buying," and "fuel problems") than the general public. Finally, the lockdown policy had significant associations with public mental health in 4 out of the 12 states we studied, among which Pennsylvania showed a positive association, whereas Michigan, North Carolina, and Ohio showed the opposite (all $P<.05$).

Conclusions: The impact of COVID-19 and the corresponding control measures on the public's mental status is dynamic and shows variability among different cohorts regarding disease types, occupations, and regional groups. Health agencies and policy makers should primarily focus on depression (reported by 51.3% of the tweets) and insomnia (which has had an ever-increasing

trend since the beginning of the pandemic), especially among health care workers. Our pipeline timely tracks and analyzes public mental health changes, especially when primary studies and large-scale surveys are difficult to conduct.

(J Med Internet Res 2022;24(10):e39676) doi:10.2196/39676

KEYWORDS

COVID-19; mental health; social media; Twitter; topic model; health care workers

Introduction

The global COVID-19 pandemic has drastically changed people's daily lives since the first confirmed case in December 2019 [1]. It has led to high hospitalization and fatality and negatively impacted public mental health [2,3]. Mental health problems cover a wide range of populations during the pandemic. The causes include but are not limited to the infection and death of relatives and friends, fear of illness, isolation brought by quarantine [4,5], and stress from unemployment [6]. At the same time, specific subpopulations such as children and adolescents [7,8], students [9,10], patients with COVID-19 [11], and health care workers [12,13] are particularly vulnerable to psychological disorders during the pandemic.

Studies have pointed out that health care workers in the United States experience psychological distress, facing high levels of anxiety, depression, and burnout during the pandemic [14]. The underlying reasons could be higher exposure risks to the virus and overwhelming workload [15,16]. Although there is literature on studying the mental health status of health care workers during the pandemic period, existing research primarily focuses on retrospective cross-sectional studies [13,14,16-19]. Therefore, it is necessary to study the dynamic characteristics of their mental status, identify general concerns, and provide timely support [20,21].

Due to their large scale, immediacy, and comprehensive coverage, social media platforms (such as Twitter, Facebook, and Weibo) have been vital data sources of research to analyze public perceptions timely when primary studies and large-scale surveys are difficult to be conducted. For example, Chew et al [22] used Twitter to study misinformation during the 2009 H1N1 pandemic, and Masri et al [23] found that new case trends can be predicted 1 week ahead based on related tweets for the 2015 Zika epidemic. Similarly, numerous studies have used social media to monitor public perceptions on topics such as enforced remote work [24], vaccines [25,26], drug use [27], mask wearing [28], and so on. Meanwhile, Berry et al [29] pointed out through a study with both quantitative and qualitative approaches that people are willing to discuss mental health problems on Twitter for varied reasons, including the sense of community and Twitter being a safe space for expression, coping, empowerment, etc. However, existing literature on public mental health during the pandemic using Twitter data [30-33] either has short study periods and small sample sizes or does not focus on subtypes of mental health problems and subgroup prevalence. More granular study designs and more comprehensive data are needed for such studies.

Finally, there is inconsistency in studying the effect of lockdown policies—one of the most highly debated topics related to mental health during the pandemic. Das et al [34] found that "state lockdown policies precede greater mental health symptoms." In contrast, Adams-Prassl et al [35] found that "the lockdown measures lowered mental health by 0.083 standard deviations."

To fill in these research gaps and potentially resolve the inconsistency, this study aimed to use related data from February 1, 2020—the beginning of the pandemic—to April 30, 2022, to analyze public mental status, problem types, their temporal and geographic distributions during COVID-19, as well as the effects of lockdown policies on public mental health across states (Figure S1 in Multimedia Appendix 1). In detail, we used this study to answer the 4 following research questions:

1. What types of mental health problems were the most frequent?
2. What mental health–related topics were the public the most concerned about, and how did relevant discussions change over time?
3. Are there differences in mental health concerns between the general population and health care workers?
4. How did lockdown policies impact public mental health?

To answer question 1, two mental health experts from our teams curated a mental health lexicon for Twitter that categorizes related tweets into 4 common mental health problems: anxiety, depression, insomnia, and addiction. Based on this lexicon, we extracted related tweets and visualized their distributions by week and state. To answer questions 2 and 3, we built a pipeline to identify potential health care workers, used a topic model to summarize related tweets into 16 topics, and compared the topic distributions among health care workers and the general population. To answer question 4, we identified tweets related to mental issues and compared their proportions before and after lockdown policies across different US states.

Methods

Data Collection

We collected and downloaded COVID-19–related tweets from February 1, 2020, to April 30, 2022, from Twitter's application programming interface using the unique tweet ID provided by an open-source COVID-19 tweet database [36]. The downloaded data contained full tweet texts and the corresponding metadata, including created time, user information, tweet status, etc. We further filtered out non–English-language and retweeted tweets and kept 471,371,477 tweets. Our data collection process strictly followed Twitter's privacy and data use management. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines.

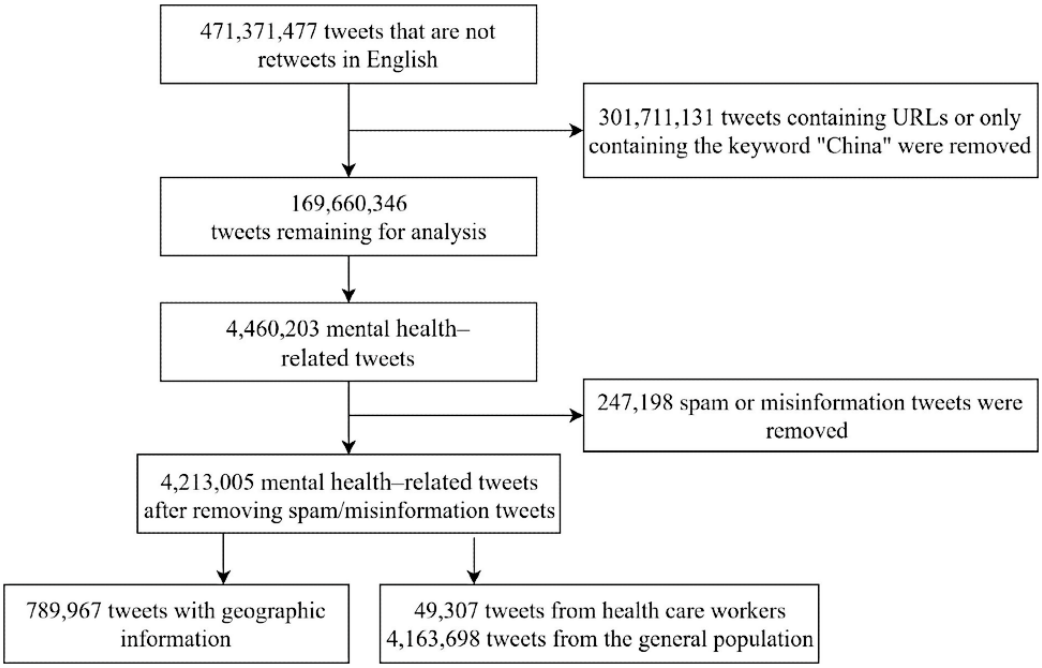
Ethics Approval

This study was conducted with approval by the Institutional Review Board of Zhejiang University (ZGL202201-2).

Data Preprocessing and Filtering

We removed tweets that contain URLs because such tweets often only included summaries or quotations of the original contents (169,660,346 tweets remained). A psychiatrist and a psychologist curated a mental health lexicon with 231 keywords. The keywords were categorized into 4 subgroups: anxiety, depression, insomnia, and addiction (Table S1 in Multimedia Appendix 1). We used this lexicon to extract mental health–related tweets through keyword matching against the preprocessed tweets and identified 4,460,203 tweets. To reduce the impact of spam and misinformation tweets, we removed data from users who posted more than 1000 mental health–related tweets during the study period. The final data set contained 4,213,005 tweets. Figure 1 shows an overview of the data preprocessing process.

Figure 1. Data collection and preprocessing.



Geographic Information Extraction

The geographic information of users was collected from 2 fields of the tweets: (1) the "place" field in tweet metadata and (2) the "location" variable nested in the "user" field of tweet metadata. The "place" information was chosen as the primary evidence of the users' geographic information, since it is generated from GPS data and is, therefore, more accurate than the information from the self-reported "location" field. We used a list of US state names to extract users' geographic information ("Methods" in Multimedia Appendix 1 [37-39]). Tweets from users associated with more than 1 state were removed in this step.

Topic Model Analysis

The Latent Dirichlet Allocation model [39] was used to conclude the main topics of mental health–related tweets. To create the corpora for topic modeling, we removed all stop words [40] as well as numbers and symbols. The topic model was implemented using the *LdaModel* function of the *Genism* package [40]. We selected the number of topics—a model hyperparameter—based on perplexity and topic coherence ("Methods" in Multimedia Appendix 1 [37-39]).

Health Care Worker Identification

To identify health care workers, we built a health care worker identification lexicon, whose keywords can be roughly divided into 3 groups: occupation, degree, and the title of the association ("Methods" in Multimedia Appendix 1 [37-39]). The dictionary contained 47 keywords, such as "doctor," "MD," "Doctor of Medicine," "FACP," etc (Table S2 in Multimedia Appendix 1). We used this lexicon to filter the user's description and extracted 49,307 tweets from health care workers.

Statistical Analysis

We applied standard descriptive statistics to summarize the 4 types of mental health–related tweets proportion, including median and IQRs. Wilcoxon matched-pairs signed-ranks test was used to compare differences between health care workers and the general population. Interrupted time series analysis [41] was applied to analyze the lockdown policy's effects on public mental health (see detailed information in "Methods" in Multimedia Appendix 1 [37-39]). We used Python software (version 3.8) to conduct the statistical analyses and chose a *P* value of .05 as the statistically significant threshold.

Results

Collected Data Set

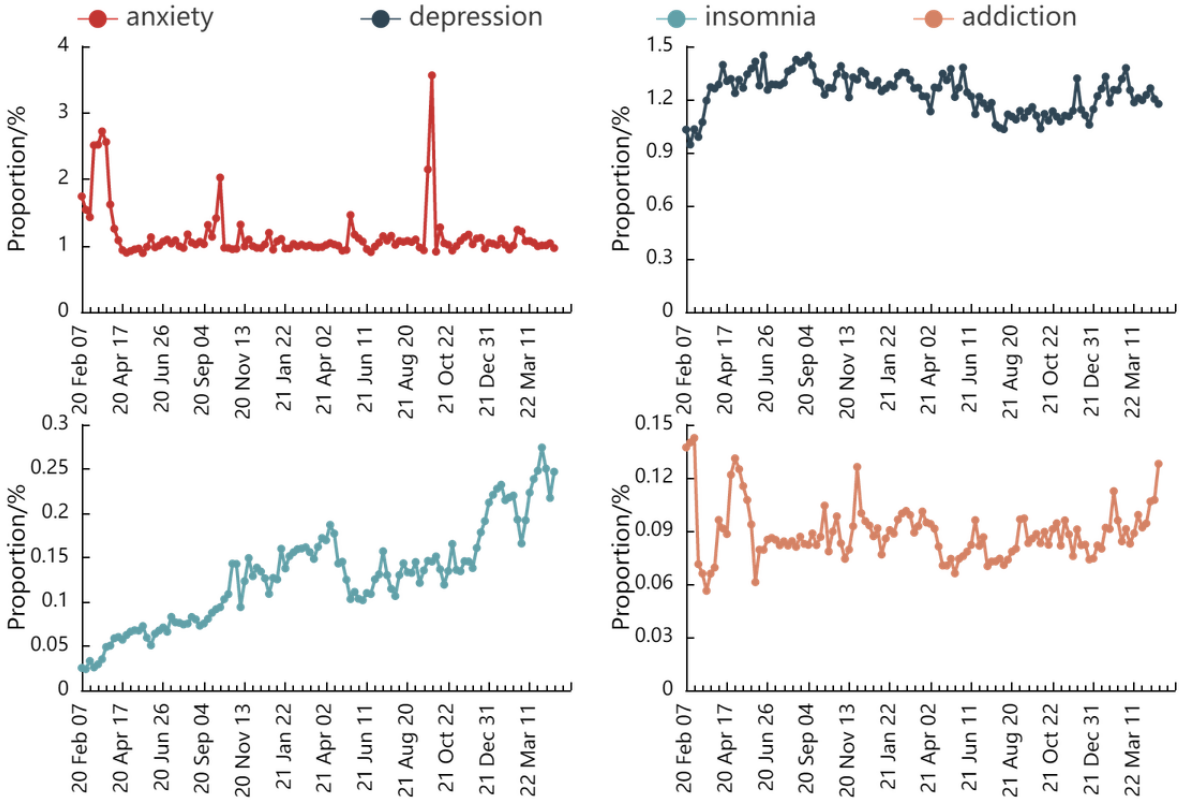
Data preprocessing selected 4,213,005 mental health–related tweets from 2,316,817 users (Figure 1). Among these tweets, 51.3% (2,161,357) were in the "depression" group, 45.66% (n=1,923,635) tweets were in the "anxiety" group, 5.35%

(n=225,205) tweets were in the "insomnia" group, and 3.56% (n=150,006) tweets were in the "addiction" group. The sum of the 4 proportions was larger than 100% because some tweets included multiple keywords that belong to different mental health subgroups. Additionally, 789,967 (18.75%) tweets were extracted with their geographic information, and health care workers posted 49,307 (1.17%) tweets (from 21,963 users).

Temporal Distribution of Mental Health–Related Tweets

The trends of the weekly numbers of COVID-19 new cases and mental health–related tweets in 4 subgroups are shown in Figure S2 in Multimedia Appendix 1. The number of tweets of mental health problems reached their first peak from February 29 to April 4, 2020. We calculated and visualized the proportions of mental health–related tweets among all COVID-19–related tweets in Figure 2. The proportion curve of anxiety-related tweets had 3 dominant peaks in March 2020, October 2020, and September 2021. The curve of insomnia-related tweets continually increased during the study period, whereas no specific trends were observed in the curves of depression and addiction.

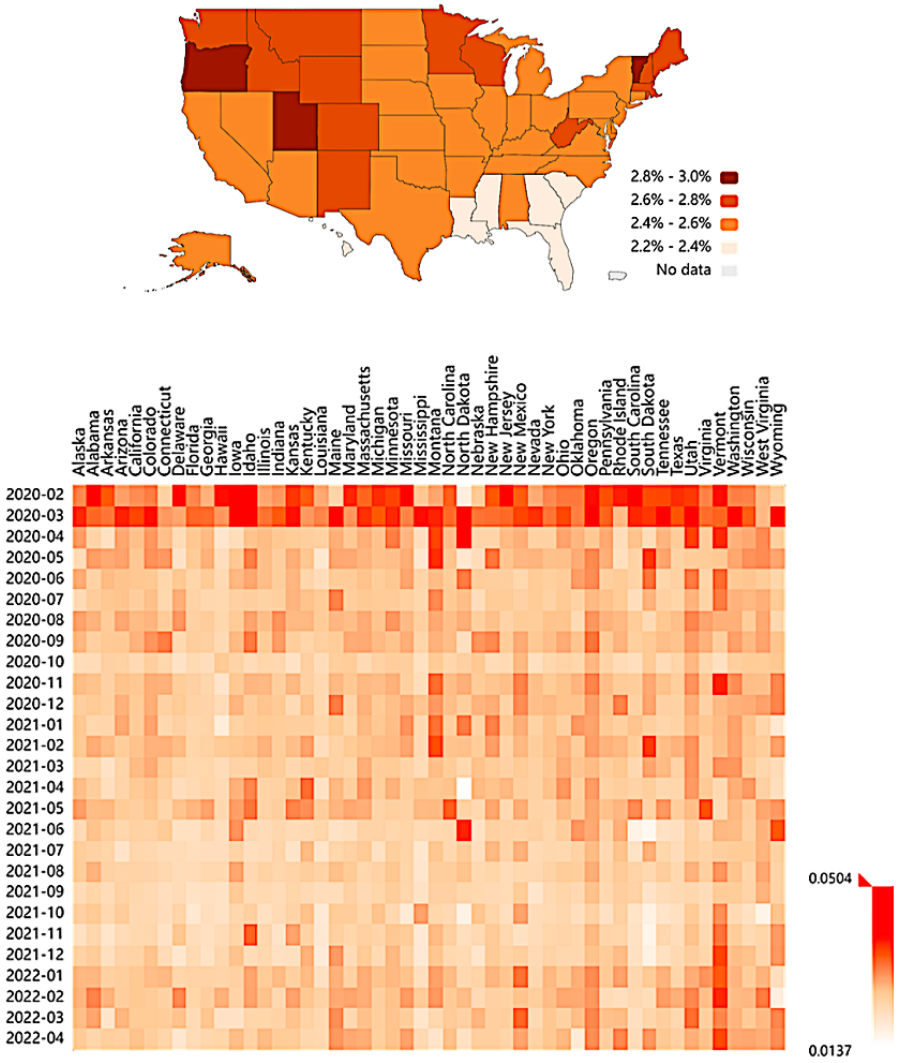
Figure 2. Trends of 4 types of mental health symptom–related tweets by the proportion of tweets.



Geographic Distribution of Mental Health–Related Tweets in the United States

Figure 3 shows the proportion of mental health–related tweets among all COVID-19–related tweets in each US state from February 1, 2020, to April 30, 2022, and visualizes the monthly tweet proportion for all the 50 US states (concrete proportions and 95% CIs are listed in Multimedia Appendix 2). Vermont, Oregon, and Utah were the 3 states with the highest proportions of mental health–related tweets, whereas Mississippi, Hawaii, and Louisiana had the lowest proportions. The first 2 months had a more substantial proportion of mental health–related tweets than the following months across most states.

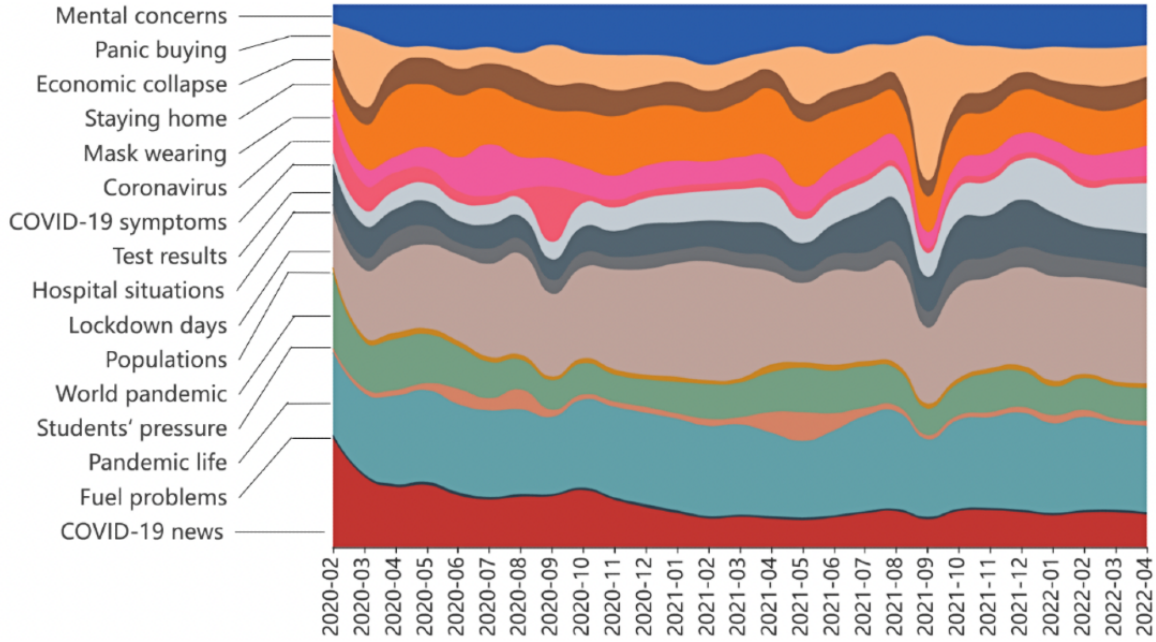
Figure 3. Proportion distribution of mental health–related tweets in the United States.



Topics of Mental Health–Related Tweets

The most frequent terms for mental health–related tweets were “people,” “worried,” “shame,” “panic,” “lockdown,” “anxiety,” “mask,” etc (Figure S3 in Multimedia Appendix 1). We chose 16 to be the number of topics based on the perplexity and coherence (“Methods” and Figure S4 in Multimedia Appendix 1 [37-39]). Topics and the corresponding top 20 most probable unigrams and bigrams are displayed in Table S3 in Multimedia Appendix 1. We assigned each topic with a topic name based on the keywords. For example, a topic having the keywords “college,” “student,” “stress,” and “exam” indicates that tweets on this topic was likely to have been focused on “students’ pressure.” Except for the issues related to COVID-19 itself, such as “COVID-19 news,” “test results,” and “mask wearing,” the public also showed particular interest in topics such as “economic collapse,” “panic buying,” and “fuel problems.” The 16 topics were then categorized into 6 topic groups: “COVID-19 pandemic,” “preventive measures,” “economic,” “people,” “education,” and “mental health.” Figure 4 shows the dynamic distributions of the investigated topics in relative tweet proportions. The topic “lockdown days” occupied a dominant position during the pandemic most of the time. “COVID-19 news” was frequently mentioned at the beginning of the pandemic but returned to an average level after June 2020. The topic of “panic buying” notably fluctuated in the research period and was relatively large from February to March 2020 and from August to October 2021.

Figure 4. Dynamic characteristics of topic proportions.



Mental Health of Health Care Workers

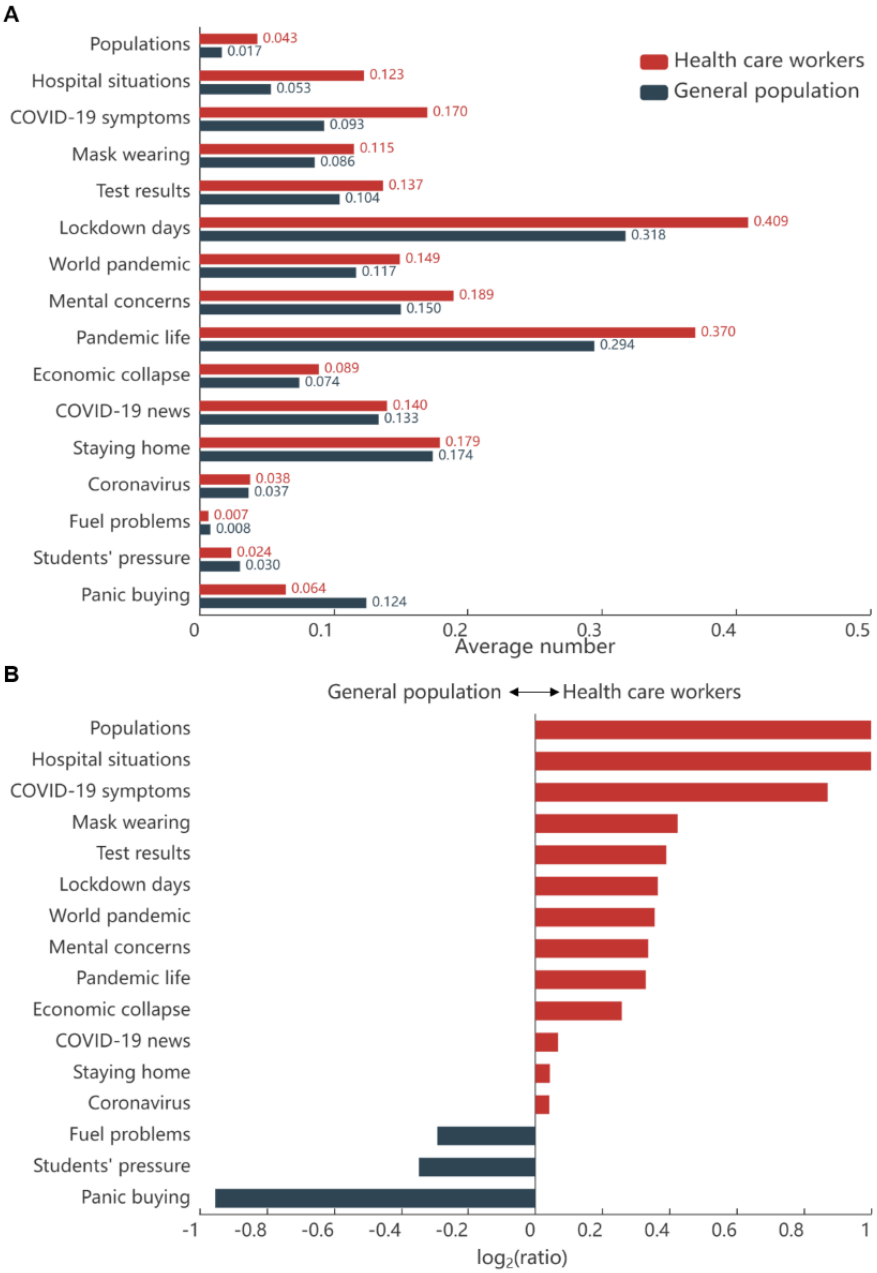
We assessed the differences in the proportions of 4 mental health symptom–related tweets between health care workers and the general population and showed the results in Table 1. Statistical results showed that the proportions of anxiety-, depression-, insomnia-, and addiction-related tweets were significantly higher in health care workers than in the general public (all P<.001). Figure 5A shows the average number of tweets per user on different topics. "Lockdown days" is the top topic discussed by both health care workers and the general population. To visualize the difference in topic distribution between health care workers and the general population, we visualized the ratios of the average number of tweets by topic for the 2 groups in Figure 5B. It demonstrates that health care workers discussed more on 13 topics, especially clinical-related topics such as "hospital situations," "COVID-19 symptoms," and "mask wearing." Conversely, the general population focused on topics such as "fuel problems," "students' pressure," and "panic buying."

Table 1. Comparison of proportions of mental health–related tweets between health care workers and the general population.

Mental health symptom	Health care workers (% tweets), median (IQR[a])	General population (% tweets), median (IQR[a])	W	*P* value
Anxiety	1.103 (1.02-1.187)	1.025 (0.956-1.094)	2120	<.001
Depression	1.519 (1.396-1.642)	1.255 (1.171-1.339)	26	<.001
Insomnia	0.251 (0.175-0.328)	0.131 (0.093-0.17)	7	<.001
Addiction	0.139 (0.114-0.164)	0.086 (0.079-0.094)	185	<.001

[a]IQR and Wilcoxon matched-pairs signed-ranks test were applied to compare the differences between the 2 groups.

Figure 5. The distribution of tweets in topics for health care workers and the general population. (A) Average number of tweets per user in each topic. (B) Logarithmic ratio of the average number of tweets between health care workers and the general population on each topic. The ratio equals the average number of tweets per user among health care workers divided by the average number of tweets among the general population.



Impacts of Lockdown Policies

We selected 12 states with more than 20,000 related tweets during the study period to explore the effect of lockdown policies on public mental status. We report the significant results found in Michigan, Pennsylvania, North Carolina, and Ohio (analysis results of the other 8 states are displayed in Figure S5 in Multimedia Appendix 1). Sensitivity analysis was applied to verify the stability of the results (Table S4 in Multimedia Appendix 1). Figure 6 shows the proportions of the 4 mental health–related tweets changed after the lockdown policy in Pennsylvania but not in the other 3 states. Table 2 lists the results of the interrupted time series analyses [41] of the lockdown policy on public mental health. The coefficient of "policy," meaning the change of intercept, was significant in the model of Pennsylvania (P=.007), and the coefficient of interaction term indicated that the change of slope was both significant in the models of Michigan (P=.03) and Pennsylvania (P=.04).

Figure 6. Daily proportion of mental health–related tweets before and after lockdown policies.

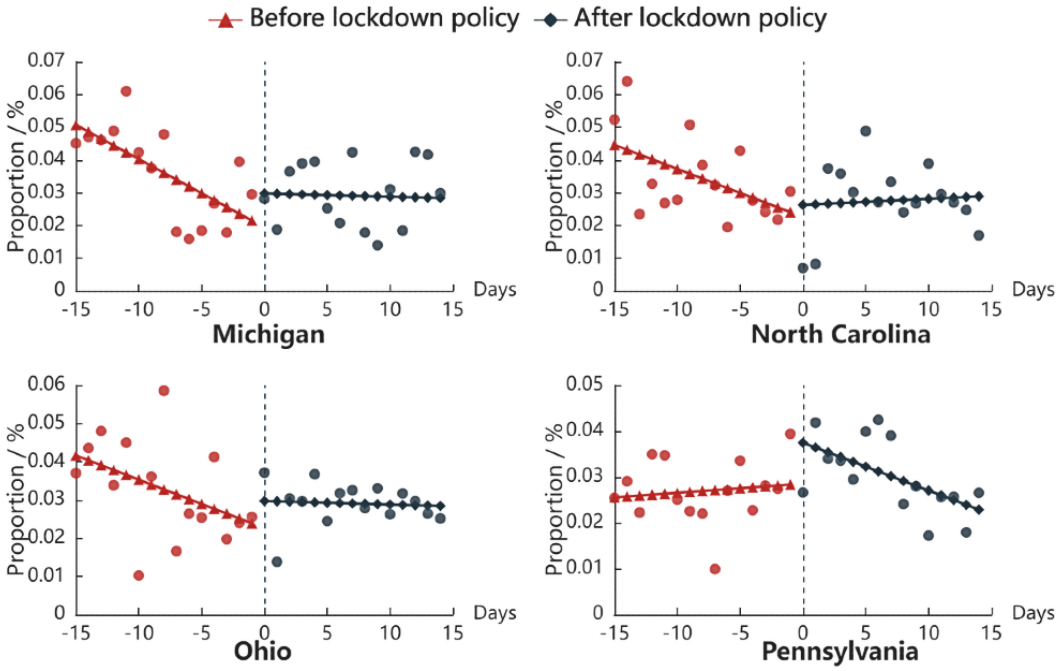


Table 2. The impact of lockdown policies on public mental health.

State	Date	Intercept	*P* value	Time[a]	*P* value	Policy[b]	*P* value	Time*policy[c]	*P* value	*F* statistic	*P* value
Michigan	March 24, 2020	0.0528	<.001	–0.0021	.003	–0.0214	.17	0.002	.03	4.669	.009
North Carolina	March 30, 2020	0.0461	<.001	–0.0015	.04	–0.0228	.16	0.0017	.08	2.509	.08
Ohio	March 23, 2020	0.0429	<.001	–0.0013	.03	–0.0117	.39	0.0012	.14	2.078	.13
Pennsylvania	April 1, 2020	0.0254	<.001	0.0002	.63	0.0288	.007	–0.0012	.04	3.033	.046

[a]Time: a continuous variable encoding the number of days in the research period (15 days before and after lockdown).

[b]Policy: a binary variable, encoded as 0 before the lockdown policy and 1 after the policy.

[c]Time*policy: the interaction term of time and policy.

Discussion

Principal Findings

We investigated public mental status for 2 and a half years since the beginning of the pandemic by analyzing topics of Twitter discussions, examining potential differences between health care workers and the general population, and studying the impacts of statewide lockdown policies. We found that anxiety and depression problems were frequently mentioned on Twitter during the study period, and the proportion of insomnia discussions increased continuously. The content analysis of mental health–related tweets revealed potential reasons: control measures, economic collapse, pressure from unemployment, and so on. Based on Twitter mentions, we found that all 4 mental health problems studied in this paper (addiction, anxiety, depression, and insomnia) were significantly more prevalent among health care workers than the general population. Finally, lockdown policies had different influences on public mental health status in different states. Among the 12 states studied, the negative effect of lockdown policies on public mental health was significant in Pennsylvania but not the other states.

Comparison to Prior Works

Consistent with research on similar topics, we found that COVID-19 has severely impacted public mental health and has dynamic influences on public mental health [30,42]. In addition, we found that the proportion of anxiety-related tweets increased to a substantial peak in March 2020 and remained low but stable for several months. A possible explanation is that the outbreak of COVID-19 caused various social problems, such as the shortage of necessities and unemployment, in the initial stage. These problems raised an intense but temporal public fear. As the pandemic continued, public concerns fell to normal as the early-stage issues were mitigated. Another possible explanation is that public emotional response diminishes as the pandemic intensifies, which is consistent with findings from Dyer and Kolic [43]. The remaining 2 peaks of anxiety-related tweets occurred during the presidential election (November 2020) and the fuel price surge (September 2021). The proportion of

insomnia also increased during the study period. This observation is consistent with Shi et al [44], who reported an incremental prevalence of insomnia in the follow-up period (from July 8 to August 8, 2020) than the baseline period (from February 28 to March 11, 2020).

The topic analysis shows that the public was concerned about the pandemic, its prevention, and the economic and educational problems caused by COVID-19. Topics such as "social distancing," "test results," "world pandemic," "COVID-19 news," and "economic collapse" were both observed in our work and previous studies [32,45-49], which only analyzed tweets during the early stage of the pandemic (mainly from January to August 2020). Our study found 2 additional topics through a longer study period: "fuel problems" and "students' pressure." These topics correspond to the literature and observations: students (especially children and adolescents) are more vulnerable to psychological disorders [50], and fuel prices frequently fluctuated during COVID-19 [51].

Unlike previous studies that only compare the prevalence of mental health symptoms between health care workers and the general population [52], we also analyzed the topics they focused on. We confirmed that health care workers were more concerned by all the studied mental problems: anxiety, depression, insomnia, and addiction. Particularly, higher proportions of insomnia among health care workers have been extensively reported in the literature [53-57]. These increased problems may be attributed to higher risks of infection [15] and more intense environmental pressure (eg, increased workload, lack of medical supplies, etc) that they face. Health care professionals were more focused on discussing the virus and more interested in sharing news or experiences related to the pandemic, demonstrating a high level of concern about the pandemic, which may be associated with an increased rate of mental disorders.

Lockdown policies had various effects on mental health discussions across US states. In Pennsylvania, it showed a positive association with mental health discussions. However, an opposite association was observed in Michigan, North Carolina, and Ohio. The literature also suggests geographically different associations between local lockdown policies and public mental health. For example, Mittal et al [58] found that most Twitter users shared positive opinions toward lockdown policies in related tweets from March 22 to April 6, 2020, whereas another study focusing on Twitter users in Massachusetts found increased anxiety expression after the enforcement of the Massachusetts State of Emergency and US State of Emergency [59]. Notably, Wang et al [60] found that public sentiment toward lockdown policies was positive in most states (such as Michigan, North Carolina, and Pennsylvania) and negative in only a few states, including Ohio, which also demonstrates geographic variations of public reactions to lockdown policies.

Strengths and Limitations

Previous work on the same topic has either not focused on the subtypes of mental health problems or studied them over short periods. Our work fills these research gaps by focusing on more granular types of mental health problems over a more extended study period. We built a comprehensive pipeline, including temporal, geographic, and discussion topic analyses; comparisons of trends and topics of concern between groups; and the impact of lockdown policies. On top of the analyses, we released the code and contributed 2 lexicons that can be used to identify mental health issues and health care professionals from tweets.

We also acknowledge the following limitations. First, the evaluation of public mental health on social media is inevitably biased due to the underlying population distribution of social media users. For example, older adults and people with low socioeconomic status may have less access to social media. As a result, this study may not reflect accurate attributes of such subpopulations. However, given the sheer number of people on Twitter, the results of this study are helpful and valuable in tracking public mental health during the pandemic. Additionally, future work could consider sampling according to users' age to avoid this problem. Second, professional psychologists must make precise diagnoses of mental health problems following official heuristics. Therefore, identifying patients using lexicons based on their tweets can introduce false cases. To validate the reliability of the lexicon, we had professional psychiatrists curate the lexicon based on sampled tweets. Third, tweets that contain keywords do not always reflect the user's mental health status as they can instead be comments on the news or from other people. To reduce this noise, we removed tweets containing URLs in our preprocessing step, as these tweets were usually summarizations or quotes of different information sources.

Future Work

The proposed pipeline can be applied to study other public mental health problems, such as suicidal thoughts, posttraumatic stress disorder, paranoia, and so on. It can also be applied to studying characteristics of other cohorts, such as sex minority groups, college students, etc. Regarding the analyses, more data sources (eg, surveys and interviews) could be introduced to validate the conclusions of this research.

Conclusions

This study developed a comprehensive pipeline to use social media for tracking and analyzing public mental status during a pandemic. It also contributed 2 lexicons that could be used in future studies. We found that the impact of COVID-19 and the corresponding control measures on the public's mental status is dynamic and shows variability among different cohorts regarding disease types, occupations, and regional groups. Health agencies and policy makers should primarily focus on depression (reported by 51.3% of the tweets) and insomnia (which has had an ever-increasing trend since the beginning of the pandemic), especially among health care workers. Our approach works efficiently, especially when primary studies and large-scale surveys are difficult to conduct. It can be extended to track the mental status of other cohorts (eg, sex minority groups and adolescents) or during different pandemic periods.

Acknowledgments

JY was partially supported by the Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province (2020E10004). The funders had no role in the design and conduct of the study.

Data Availability

The data and code supporting the study's findings are available at https://github.com/zjumh/mental-health-during-COVID.

Authors' Contributions

ML and JY designed the study and drafted the manuscript. YH prepared the data, provided feedback on the study design, and helped draft and revise the manuscript. ML performed data and statistical analysis. YL and LW built the lexicon of mental health keywords. YL, LZ, and XL provided critical reviews. All authors reviewed the manuscript. ML takes responsibility for the integrity of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods, pictures, and tables.
[DOCX File , 1138 KB - jmir_v24i10e39676_app1.docx]

Multimedia Appendix 2
The proportion and 95% CIs of mental health–related tweets in each state by month.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v24i10e39676_app2.xlsx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
3. Vigo D, Patten S, Pajer K, Krausz M, Taylor S, Rush B, et al. Mental health of communities during the COVID-19 pandemic. Can J Psychiatry 2020 Oct;65(10):681-687 [FREE Full text] [doi: 10.1177/0706743720926676] [Medline: 32391720]
4. Wang Y, Shi L, Que J, Lu Q, Liu L, Lu Z, et al. The impact of quarantine on mental health status among general population in China during the COVID-19 pandemic. Mol Psychiatry 2021 Sep;26(9):4813-4822 [FREE Full text] [doi: 10.1038/s41380-021-01019-y] [Medline: 33483692]
5. Xin M, Luo S, She R, Yu Y, Li L, Wang S, et al. Negative cognitive and psychological correlates of mandatory quarantine during the initial COVID-19 outbreak in China. Am Psychol 2020;75(5):607-617. [doi: 10.1037/amp0000692] [Medline: 32673008]
6. Witteveen D, Velthorst E. Economic hardship and mental health complaints during COVID-19. Proc Natl Acad Sci U S A 2020 Nov 03;117(44):27277-27284 [FREE Full text] [doi: 10.1073/pnas.2009609117] [Medline: 33046648]
7. Golberstein E, Wen H, Miller BF. Coronavirus disease 2019 (COVID-19) and mental health for children and adolescents. JAMA Pediatr 2020 Sep 01;174(9):819-820. [doi: 10.1001/jamapediatrics.2020.1456] [Medline: 32286618]
8. Leeb RT, Bitsko RH, Radhakrishnan L, Martinez P, Njai R, Holland KM. MMWR Morb Mortal Wkly Rep 2020 Nov 13;69(45):1675-1680 [FREE Full text] [doi: 10.15585/mmwr.mm6945a3] [Medline: 33180751]
9. Copeland WE, McGinnis E, Bai Y, Adams Z, Nardone H, Devadanam V, et al. Impact of COVID-19 pandemic on college student mental health and wellness. J Am Acad Child Adolesc Psychiatry 2021 Jan;60(1):134-141.e2 [FREE Full text] [doi: 10.1016/j.jaac.2020.08.466] [Medline: 33091568]
10. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020 Sep 17;22(9):e22817 [FREE Full text] [doi: 10.2196/22817] [Medline: 32897868]
11. Naidu SB, Shah AJ, Saigal A, Smith C, Brill SE, Goldring J, et al. The high mental health burden of "Long COVID" and its association with on-going physical and respiratory symptoms in all adults discharged from hospital. Eur Respir J 2021 Jun;57(6):2004364 [FREE Full text] [doi: 10.1183/13993003.04364-2020] [Medline: 33795319]
12. Lin K, Yang BX, Luo D, Liu Q, Ma S, Huang R, et al. The mental health effects of COVID-19 on health care providers in China. Am J Psychiatry 2020 Jul 01;177(7):635-636. [doi: 10.1176/appi.ajp.2020.20040374] [Medline: 32605443]

13. Bryant-Genevier J, Rao CY, Lopes-Cardozo B, Kone A, Rose C, Thomas I, et al. Symptoms of depression, anxiety, post-traumatic stress disorder, and suicidal ideation among state, tribal, local, and territorial public health workers during the COVID-19 pandemic - United States, March-April 2021. MMWR Morb Mortal Wkly Rep 2021 Dec 03;70(48):1680-1685 [FREE Full text] [doi: 10.15585/mmwr.mm7048a6] [Medline: 34855723]
14. Firew T, Sano ED, Lee JW, Flores S, Lang K, Salman K, et al. Protecting the front line: a cross-sectional survey analysis of the occupational factors contributing to healthcare workers' infection and psychological distress during the COVID-19 pandemic in the USA. BMJ Open 2020 Oct 21;10(10):e042752 [FREE Full text] [doi: 10.1136/bmjopen-2020-042752] [Medline: 33087382]
15. Rudberg AS, Havervall S, Månberg A, Jernbom Falk A, Aguilera K, Ng H, et al. SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden. Nat Commun 2020 Oct 08;11(1):5064 [FREE Full text] [doi: 10.1038/s41467-020-18848-0] [Medline: 33033249]
16. Manzano García G, Ayala Calvo JC. The threat of COVID-19 and its influence on nursing staff burnout. J Adv Nurs 2021 Feb;77(2):832-844. [doi: 10.1111/jan.14642] [Medline: 33155716]
17. Amin F, Sharif S, Saeed R, Durrani N, Jilani D. COVID-19 pandemic- knowledge, perception, anxiety and depression among frontline doctors of Pakistan. BMC Psychiatry 2020 Sep 23;20(1):459 [FREE Full text] [doi: 10.1186/s12888-020-02864-x] [Medline: 32967647]
18. Chew NWS, Lee GKH, Tan BYQ, Jing M, Goh Y, Ngiam NJH, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
19. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
20. Rimmer A. COVID-19: drop the hero narrative and support doctors' mental health, says charity. BMJ 2021 Feb 04;372:n337. [doi: 10.1136/bmj.n337] [Medline: 33541858]
21. Rimmer A. COVID-19: Two fifths of doctors say pandemic has worsened their mental health. BMJ 2020 Oct 27;371:m4148. [doi: 10.1136/bmj.m4148] [Medline: 33109530]
22. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
23. Masri S, Jia J, Li C, Zhou G, Lee M, Yan G, et al. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. BMC Public Health 2019 Jun 14;19(1):761 [FREE Full text] [doi: 10.1186/s12889-019-7103-8] [Medline: 31200692]
24. Zhang C, Yu MC, Marin S. Exploring public sentiment on enforced remote work during COVID-19. J Appl Psychol 2021 Jun;106(6):797-810. [doi: 10.1037/apl0000933] [Medline: 34138587]
25. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, et al. Public perception of COVID-19 vaccines on Twitter in the United States. medRxiv Preprint posted online on October 18, 2021 [FREE Full text] [doi: 10.1101/2021.10.16.21265097] [Medline: 34704100]
26. 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]
27. Hua Y, Jiang H, Lin S, Yang J, Plasek J, Bates D, et al. Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. J Am Med Inform Assoc 2022 Sep 12;29(10):1668-1678 [FREE Full text] [doi: 10.1093/jamia/ocac114] [Medline: 35775946]
28. Sanders AC, White RC, Severson LS, Ma R, McQueen R, Alcântara Paulo HC, et al. Unmasking the conversation on masks: natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. AMIA Jt Summits Transl Sci Proc 2021;2021:555-564 [FREE Full text] [Medline: 34457171]
29. 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]
30. Zhang S, Sun L, Zhang D, Li P, Liu Y, Anand A, et al. The COVID-19 pandemic and mental health concerns on Twitter in the United States. Health Data Sci 2022 Feb 17;2022:1-9. [doi: 10.34133/2022/9758408]
31. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on Twitter during the COVID-19 pandemic: observational study. JMIR Infodemiology 2021 Jul 16;1(1):e26769 [FREE Full text] [doi: 10.2196/26769] [Medline: 34458682]
32. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
33. Koh JX, Liew TM. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds. J Psychiatr Res 2022 Jan;145:317-324 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.015] [Medline: 33190839]

34. Das A, Singh P, Bruckner TA. State lockdown policies, mental health symptoms, and using substances. Addict Behav 2022 Jan;124:107084 [FREE Full text] [doi: 10.1016/j.addbeh.2021.107084] [Medline: 34507184]
35. Adams-Prassl A, Boneva T, Golin M, Rauh C. The impact of the coronavirus lockdown on mental healthvidence from the United States. Econ Policy 2022 Jan 09;37(109):139-155. [doi: 10.1093/epolic/eiac002]
36. 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]
37. Newman D, Lau JH, Grieser K, Baldwin T. Automatic evaluation of topic coherence. 2010 Jun Presented at: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics; June 2-4, 2010; Los Angeles, CA p. 100-108 URL: https://aclanthology.org/N10-1012
38. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017 Feb 01;46(1):348-355 [FREE Full text] [doi: 10.1093/ije/dyw098] [Medline: 27283160]
39. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan 03;3:993-1022 [FREE Full text]
40. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. 2010 Presented at: LREC 2010: New Challenges for NLP Frameworks; May 22, 2010; Valletta, Malta p. 46-50 URL: https://www.fi.muni.cz/usr/sojka/papers/lrec2010-rehurek-sojka.pdf
41. Biglan A, Ary D, Wagenaar A. The value of interrupted time-series experiments for community intervention research. Prev Sci 2000 Mar;1(1):31-49 [FREE Full text] [doi: 10.1023/a:1010024016308] [Medline: 11507793]
42. 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]
43. Dyer J, Kolic B. Public risk perception and emotion on Twitter during the COVID-19 pandemic. Appl Netw Sci 2020 Dec 16;5(1):99 [FREE Full text] [doi: 10.1007/s41109-020-00334-7] [Medline: 33344760]
44. Shi L, Lu Z, Que J, Huang X, Lu Q, Liu L, et al. Long-term impact of COVID-19 on mental health among the general public: a nationwide longitudinal study in China. Int J Environ Res Public Health 2021 Aug 20;18(16):8790 [FREE Full text] [doi: 10.3390/ijerph18168790] [Medline: 34444539]
45. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
46. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
47. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
48. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
49. Kwon S, Park A. Understanding user responses to the COVID-19 pandemic on Twitter from a terror management theory perspective: cultural differences among the US, UK and India. Comput Human Behav 2022 Mar;128:107087 [FREE Full text] [doi: 10.1016/j.chb.2021.107087] [Medline: 34744298]
50. Singh S, Roy D, Sinha K, Parveen S, Sharma G, Joshi G. Impact of COVID-19 and lockdown on mental health of children and adolescents: a narrative review with recommendations. Psychiatry Res 2020 Nov;293:113429 [FREE Full text] [doi: 10.1016/j.psychres.2020.113429] [Medline: 32882598]
51. Weekly retail gasoline and diesel prices. U.S. Energy Information Administration. URL: https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_w.htm [accessed 2022-10-05]
52. Sakib N, Akter T, Zohra F, Bhuiyan AKMI, Mamun MA, Griffiths MD. Fear of COVID-19 and depression: a comparative study among the general population and healthcare professionals during COVID-19 pandemic crisis in Bangladesh. Int J Ment Health Addict 2021 Feb 19:1-17 [FREE Full text] [doi: 10.1007/s11469-020-00477-9] [Medline: 33642957]
53. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
54. Li Y, Chen B, Hong Z, Sun Q, Dai Y, Basta M, et al. Insomnia symptoms during the early and late stages of the COVID-19 pandemic in China: a systematic review and meta-analysis. Sleep Med 2022 Mar;91:262-272 [FREE Full text] [doi: 10.1016/j.sleep.2021.09.014] [Medline: 34732293]
55. 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]

56. Cénat JM, Blais-Rochette C, Kokou-Kpolou CK, Noorishad P, Mukunzi JN, McIntee S, et al. Prevalence of symptoms of depression, anxiety, insomnia, posttraumatic stress disorder, and psychological distress among populations affected by the COVID-19 pandemic: a systematic review and meta-analysis. Psychiatry Res 2021 Jan;295:113599 [FREE Full text] [doi: 10.1016/j.psychres.2020.113599] [Medline: 33285346]
57. Jahrami H, BaHammam AS, Bragazzi NL, Saif Z, Faris M, Vitiello MV. Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. J Clin Sleep Med 2021 Feb 01;17(2):299-313 [FREE Full text] [doi: 10.5664/jcsm.8930] [Medline: 33108269]
58. Mittal R, Ahmed W, Mittal A, Aggarwal I. Twitter users' coping behaviors during the COVID-19 lockdown: an analysis of tweets using mixed methods. Inf Discov Deliv 2021 May 15;49(3):193-202. [doi: 10.1108/idd-08-2020-0102]
59. Thorpe Huerta D, Hawkins JB, Brownstein JS, Hswen Y. Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: a Twitter analysis. SSM Popul Health 2021 Sep;15:100851 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100851] [Medline: 34355055]
60. Wang J, Fan Y, Palacios J, Chai Y, Guetta-Jeanrenaud N, Obradovich N, et al. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. Nat Hum Behav 2022 Mar;6(3):349-358. [doi: 10.1038/s41562-022-01312-y] [Medline: 35301467]

Edited by T Leung; submitted 18.05.22; peer-reviewed by A Pal, J Ruiyu; comments to author 15.06.22; revised version received 21.07.22; accepted 30.09.22; published 13.10.22.

Please cite as:
Li M, Hua Y, Liao Y, Zhou L, Li X, Wang L, Yang J
Tracking the Impact of COVID-19 and Lockdown Policies on Public Mental Health Using Social Media: Infoveillance Study
J Med Internet Res 2022;24(10):e39676
URL: https://www.jmir.org/2022/10/e39676
doi:10.2196/39676
PMID:36191167

Viewpoint

The Post-Roe Political Landscape Demands a Morality of Caution for Women's Health

Sarah Goodday[1,2], PhD; Daniel Karlin[1,3,4], MD, MA; Christine Suver[5], PhD; Stephen Friend[1,2], MD, PhD

[1]4YouandMe, Seattle, WA, United States

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

[3]Department of Psychiatry, Tufts University School of Medicine, Boston, MA, United States

[4]MindMed Inc, New York, NY, United States

[5]Sage Bionetworks, Seatlle, WA, United States

Corresponding Author:
Sarah Goodday, PhD
4YouandMe
2901 Third Avenue
Ste 330
Seattle, WA, 98121
United States
Phone: 1 9022374235
Email: sarah@4youandme.org

Abstract

The recent Supreme Court decision (ie, Dobbs v. Jackson Women's Health Organization), revoking the constitutional right to abortion in the United States, has the potential to dramatically disrupt progress in women's health research. The typical safeguards to ensure confidentiality and privacy of research participants in studies that collect certain types of personal health information may not hold against criminal investigations surrounding suspected pregnancy terminations. There are additional risks to participants in digital health research studies involving the use of wearable devices capable of tracking physiological measures, such as body temperature and heart rate, as these have shown promise for tracking conception and could be used to identify pregnancy termination signatures. There are strategies researchers can use to protect the safety of participants in health research who could get pregnant, while also maintaining integrity of research methods. The objective of this viewpoint is to discuss potential strategies to protect research participants' privacy that include the minimization of nonessential sensitive personal health information and anonymization protocols in the event of miscarriage or termination of pregnancy. We invite others to join this discussion so as to not let the current political landscape impede progress in women's health and reproductive research, while also protecting research participants.

(J Med Internet Res 2022;24(10):e41417) doi:10.2196/41417

KEYWORDS

women's health; reproductive health; wearable; abortion rights, confidentiality and privacy; Roe v. Wade; health policy; health research; reproductive information; privacy; women's rights; health rights; abortion; eHealth; digital health; mHealth; safety; ethic

Introduction

The US Supreme Court's decision involving the case Dobbs v. Jackson Women's Health Organization (Dobbs v. Jackson ruling) in June 2022 to overturn Roe v. Wade, thereby dismissing the constitutional right to abortion [1] led to several US states taking rapid action to ban, restrict, and criminalize abortion. This decision will have significant negative impacts on maternal and child health, their economic welfare, well-being outcomes, and mortality in the United States, disproportionately impacting those from disadvantaged populations [2,3]. This decision will also dramatically disrupt progress in women's health and reproductive research.

Numerous policy efforts have been developed to enhance the inclusion in health research of women and other people who can get pregnant, especially in light of the historical exclusion of this group, including many efforts by the National Institutes of Health [4-8]. The Dobbs v. Jackson ruling dispiritedly impacts this progress, where participants in health research who can get pregnant are now at an increased risk of having certain types of personal health information used against them by some states or other individuals. Until now, investigators were able to rely

on maintaining confidentiality of their research participants through protections granted by federal or state statutes, for example, through Health Insurance Portability and Accountability Act (HIPAA) compliance, which applies restrictions on the use and disclosure of personal health information. Additional layers of protection could be added through safeguards, such as a Certificate of Confidentiality, that are issued pursuant to the 21st Century Cures Act's amendments to the Public Health Services Act [9] and allows participants and investigators to refuse the disclosure of research data in the event of a federal, state, or local request; however, both HIPAA and the Certificate of Confidentiality have exclusions that are not universally applicable in research projects in the United States. Since the Dobbs v. Jackson ruling, these governance safeguards have been called into question, suggesting that they may not be sufficient to protect participants' confidentiality in the face of a criminal investigation related to suspicion of abortion [10-12]. These recent developments have sparked discussion on additional mechanisms to protect patient privacy, especially in states that criminalize abortion and more so in those that work to restrict individuals' ability to travel out of state to access abortion clinics. As of August 2022, a total of 12 states had a full ban on abortion (ie, AL, AR, ID, KY, LA, MS, MO, OK, SD, TN, TX, and WI), and 2 states had a gestational limit of 6 weeks (GA and OH) [13].

The reversal of abortion rights in the United States demands a 'morality of caution' around the collection of personal health information in health research that includes women and other people who can get pregnant [14]. The new political landscape surrounding abortion poses an immediate risk to participants, particularly those engaged in pregnancy or reproductive health-related research, but also those engaged in general biomedical research, where certain types of personal health information collected could be used to identify pregnancy termination events (eg, GPS data). This creates a pressing challenge for health researchers and a need to find solutions into the future. The objective of this viewpoint is to outline potential solutions for researchers that offer stronger protection for participants while maintaining integrity of research methods.

The obvious yet unfortunate solution for participants to completely reduce their personal risk and inadvertently offer any information that could be used against them is to not participate in health research. Further, participants in current research could request to have their personal health information deleted, akin to the European Union's General Data Protection Regulation "Right to Be Forgotten" [15]; however, this relies on the participants having adequate knowledge of their risk as a participant, which is not always transparent, and the willingness of research institutions to honor such requests. We urge investigators and participants to consider alternative approaches so as to not impede progress in women's health and reproductive research.

Potential Solutions for Health Researchers

First, investigators should consider minimizing the information they collect, particularly sensitive information that could be used to indicate a miscarriage or pregnancy termination where this is not essential to meeting the study objectives. In these cases, researchers can consider discarding certain questions from existing surveys or questionnaires. If the data are not collected, they cannot be used to prosecute a participant. Although effective and able to preserve the data needed to meet the primary study objectives, this intervention omits data that could have added value in exploratory post hoc analyses or in integrated data sets. Therefore, in this approach, the consequences of censoring the collection of such data that are integral to health research should not be ignored.

There are two challenges to this aforementioned approach. First, open science research studies that collect reproductive health data that run without a data lock that controls access to research data—that is, data that are hosted in the public domain will be challenging to safeguard; in these open science environments, researchers may only be able to protect prospective participants through the removal of sensitive data fields before uploading the data. Second, if the information collected involves digital passive data, the removal of key sensitive fields becomes more complex. The increasing prevalence of digital health apps that are intended to provide useful information on the menstrual cycle and reproductive health, use a variety of smartphone-based tracking features, and often incorporate wearable devices poses potential risks to users. These risks are amplified by the substantial lack of regulation around the use of digital data from health-tracking apps and wearables [16], as such data are not subject to HIPAA regulations. Wearable devices capable of tracking physiological measures, such as body temperature and heart rate, have shown promise for tracking conception [17-20], and studies are starting to explore their potential for health monitoring in pregnancy [21]. These same signals could likely be used to identify miscarriage or termination signatures in study participants. This means that even without purposefully labeling miscarriage or termination events from surveys and questionnaires or health record data, a participant could still be at risk that their digital data be used to infer changes in their reproductive state or access to services. Additional passive data fields, such as GPS coordinates, activity, phone records, and many more, could also be used to determine abortion clinic access.

A second strategy for protecting research participant data from invasive investigation is for researchers to execute a strict and preferably automated anonymization protocol on any participant-level data as soon as a miscarriage or termination event occurs—that is, immediately deleting all personal identifiable data relevant to a participant who is no longer pregnant, including any keys linking study identifiers to personal identifiable data. In doing so, a substantial barrier between sensitive information and the participant is created. In light of the potential risk related to data acquired from wearable devices, this second alternative may be necessary to maximally protect participants. The consequence of this solution is that if it is executed during an active study period, those participants become untraceable and uncontactable for any follow-up study activities, resulting in inadvertent loss of other data points and potential study completion challenges. However, in the context of a research study where a pregnancy loss is a study end point,

this consequence is likely to be less impactful. Additional consideration must be taken into ensuring this approach does not inadvertently further impose inequitable data deletions that differ by sex and gender. In using the aforementioned approaches, ensuring participants are fully informed prior to their participation and offering a choice is crucial.

Finally, efforts should be enhanced to develop reliable methods to generate synthetic data and other breakthrough technologies that preserve the value of the data while obfuscating the real data. Synthetic data sets are simulated data sets that retain the structure and statistical distribution of the original data set. When accurate, these artificially created data sets could be used in analysis and modeling without revealing the real-world data. However, outliers and small data sets remain challenging to simulate in a synthetic data set.

Conclusions

As the reality of the Dobbs v. Jackson ruling sets in, we urge researchers to be proactive in activating processes and procedures to enable full engagement of women and others who can get pregnant in health research studies, while considering appropriate precautions for their privacy and safety now and in future studies. Although we have highlighted some solutions here, there are undoubtedly many other solutions that will surface as the political landscape continues to evolve. As a community, we must do everything possible to protect research participants, while also not impeding progress in reproductive and women's health research.

Conflicts of Interest

None declared.

References

1. Dobbs v. Jackson Women's Health Organization, 597 U.S. ___ (2022). URL: https://supreme.justia.com/cases/federal/us/597/19-1392/ [accessed 2022-10-11]
2. Cohen SA. Abortion and Women of Color: The Bigger Picture. 2008. URL: https://www.guttmacher.org/gpr/2008/08/abortion-and-women-color-bigger-picture [accessed 2022-10-12]
3. Gostin L, Reingold R. Ending the constitutional right to abortion in the United States. BMJ 2022 Aug 01;378:o1897 [FREE Full text] [doi: 10.1136/bmj.o1897] [Medline: 35914779]
4. Arnegard ME, Whitten LA, Hunter C, Clayton JA. Sex as a biological variable: a 5-year progress report and call to action. J Womens Health (Larchmt) 2020 Jun;29(6):858-864 [FREE Full text] [doi: 10.1089/jwh.2019.8247] [Medline: 31971851]
5. NIH guidelines on the inclusion of women and minorities as subjects in clinical research. NIH Grants & Funding. URL: https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm [accessed 2022-10-11]
6. Amendment: NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research. 2017. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html [accessed 2022-10-12]
7. Institute of Medicine (US) Committee on Understanding the Biology of Sex and Gender Differences. In: Wizemann TM, Pardue ML, editors. Exploring the Biological Contributions to Human Health: Does Sex Matter?. Washington, DC: National Academics Press (US); 2001.
8. Shafir R, Olson E, Colloca L. The neglect of sex: a call to action for including sex as a biological variable in placebo and nocebo research. Contemp Clin Trials 2022 May;116:106734. [doi: 10.1016/j.cct.2022.106734] [Medline: 35306216]
9. Wolf L, Beskow L. Certificates of confidentiality: mind the gap. Utah Law Rev 2021;2021(4):937-950 [FREE Full text] [doi: 10.26054/0d-80m8-7v82] [Medline: 35529681]
10. Boodman E, Bannow T, Herman B, Ross C. HIPAA won't protect you if prosecutors want your reproductive health records. STAT. 2022. URL: https://www.statnews.com/2022/06/24/hipaa-wont-protect-you-if-prosecutors-want-your-reproductive-health-records/ [accessed 2022-10-11]
11. HIPAA and reproductive health. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/reproductive-health/index.html [accessed 2022-10-11]
12. Cohen K. Location, health, and other sensitive information: FTC committed to fully enforcing the law against illegal use and sharing of highly sensitive data. Federal Trade Commission. 2022. URL: https://www.ftc.gov/business-guidance/blog/2022/07/location-health-other-sensitive-information-ftc-committed-fully-enforcing-law-against-illegal-use [accessed 2022-10-12]
13. Tracking the states where abortion is now banned. The New York Times. 2022. URL: https://www.nytimes.com/interactive/2022/us/abortion-laws-roe-v-wade.html [accessed 2022-10-11]
14. Matheson J. Moral caution and the epistemology of disagreement. J Soc Philos 2016 Jun 07;47(2):120-141 [FREE Full text] [doi: 10.1111/josp.12145]
15. Everything you need to know about the "Right to be Forgotten". GDPR. URL: https://gdpr.eu/right-to-be-forgotten/ [accessed 2022-10-11]
16. Knox R, Tenenbaum C. Regulating digital health apps needs user-centered reform. STAT. 2021. URL: https://www.statnews.com/2021/08/03/refor-regulatory-landscape-digital-health-applications/ [accessed 2022-10-11]

17. Grant A, Smarr B. Feasibility of continuous distal body temperature for passive, early pregnancy detection. PLOS Digit Health 2022 May 16;1(5):e0000034 [FREE Full text] [doi: 10.1371/journal.pdig.0000034]
18. Smarr BL, Zucker I, Kriegsfeld LJ. Detection of Successful and Unsuccessful Pregnancies in Mice within Hours of Pairing through Frequency Analysis of High Temporal Resolution Core Body Temperature Data. PLoS One 2016;11(7):e0160127 [FREE Full text] [doi: 10.1371/journal.pone.0160127] [Medline: 27467519]
19. Grant AD, Newman M, Kriesgfeld LJ. Ultradian rhythms in heart rate variability and distal body temperature anticipate onset of the luteinizing hormone surge. Sci Rep 2020;10(1):20378. [doi: 10.1038/s41598-020-76236-6] [Medline: 33230235]
20. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable Oura ring. BMC Womens Health 29;19 (1). pmid 2019:a.
21. Goodday SM, Karlin E, Brooks A, Chapman C, Karlin DR, Foschini L, et al. Better Understanding of the Metamorphosis of Pregnancy (BUMP): protocol for a digital feasibility study in women from preconception to postpartum. NPJ Digit Med 2022 Mar 30;5(1):40. [doi: 10.1038/s41746-022-00579-9] [Medline: 35354895]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act

Edited by G Eysenbach, T Leung, R Kukafka; submitted 25.07.22; peer-reviewed by L Dodge, K Vallury, Z Zandesh, J Wagner, J Wilbanks, N Cobb; comments to author 30.08.22; revised version received 08.09.22; accepted 07.10.22; published 20.10.22.

Please cite as:
Goodday S, Karlin D, Suver C, Friend S
The Post-Roe Political Landscape Demands a Morality of Caution for Women's Health
J Med Internet Res 2022;24(10):e41417
URL: https://www.jmir.org/2022/10/e41417
doi:10.2196/41417
PMID:36264611

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org